

This Page Is Inserted by IFW Operations
and is not a part of the Official Record

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

**As rescanning documents *will not* correct images,
please do not report the images to the
Image Problem Mailbox.**

THIS PAGE BLANK (USPTO)

GenCore version 5.1.4_p5-4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 26, 2003, 08:59:02 ; Search time 1931 Seconds

(without alignments) 14378.082 Million cell updates/sec

Title: US-09-880-729A-1

Perfect score: 954
Sequence: 1 atgggtgtgtatccttttga.....gagagatagatgataaa 954

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_da:*
2: gb_htg:*
3: gb_in:*
4: gb_cm:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vl:*
15: em_da:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vl:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rtd:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	954	100.0	954	6	AR077712	AR077712 Sequence
2	954	100.0	954	6	AR096552	AR096552 Sequence
3	954	100.0	954	6	AR157661	AR157661 Sequence
4	952.4	99.8	10244	1	AE001813	AE001813 Sequence
5	235.8	24.7	1083	6	AR205090	AR205090 Sequence
6	153.2	16.1	2160	1	CLOCCELCD	D90341 C. cellulyoly
7	151.6	15.9	3002	1	CLOCCEH	M31903 C. thermocel
8	82.2	8.6	10415	1	AE007598	AE007598 Clostridi
9	78.6	8.2	1756	8	NPE0DGEUB	Z31364 N. patriciar
10	76	8.0	3471	1	CLOCCELEX	M22759 Clostridium
11	75.4	7.9	1451	8	AF015249	AF015249 Orpinomyces
12	74.8	7.8	1876	8	AF165266	AF165266 Pyromyces
13	73	7.7	1825	8	OSU57818	U57818 Orpinomyces
14	73	7.7	1826	6	AR130363	AR130363 Sequence
15	72.6	7.6	5502	8	PE0277483	AJ277483 Pyromyces
16	71.8	7.5	1998	1	CLOCENGB	M75706 Clostridium
17	68.6	7.2	1532	6	AR108136	AR108136 Sequence
18	68.6	7.2	1532	8	U97153	U97153 Orpinomyces
19	68.6	7.2	1966	8	AF015248	AF015248 Orpinomyces
20	63.2	6.6	3949	8	AF053363	AF053363 Neocallim
21	62.8	6.6	1086	8	AF094757	AF094757 Pyromyces
22	60.4	6.3	7218	6	I66494	I66494 Sequence 14
23	58.6	6.1	2000	1	CLOCENGD	M37434 Clostridium
24	58.4	6.1	1634	8	NFU38843	U38843 Neocallim
25	58.2	6.1	10705	1	AE007666	AE007666 Clostridi
26	53.8	5.6	10894	1	AE007844	AE007844 Clostridi
27	53	5.6	247699	2	AC106241	AC106241 Rattus no
28	50.6	5.3	1550	1	RUMCELANA	L10243 Ruminooccu
29	50	5.2	163208	2	AC124923	AC124923 Rattus no
30	50	5.2	300750	1	AP001509	AP001509 Bacillus
31	47.6	5.0	149834	2	AC105683	AC105683 Rattus no
32	47.4	5.0	146383	2	AC116367	AC116367 Oryza sat
33	46.8	4.9	166096	2	AC120947	AC120947 Rattus no
34	46.6	4.9	1141	6	AX083744	AX083744 Sequence
35	45.6	4.8	64688	2	AC131425	AC131425 Rattus no
36	45.6	4.8	161844	2	AC106526	AC106526 Rattus no
37	45.2	4.7	125020	9	AF429315	AF429315 Homo sapi
38	45.2	4.7	200820	2	AC129459	AC129459 Rattus no
39	45	4.7	122737	2	AC109901	AC109901 Rattus no
40	45	4.7	124524	9	AC117411	AC117411 Homo sapi
41	44.8	4.7	774	11	CNS06FFX	AL396355 T3 end of
42	44.8	4.7	58696	2	AC103749	AC103749 Homo sapi
43	44.8	4.7	161389	10	AL691440	AL691440 Mouse DNA
44	44.8	4.7	171462	2	AC117832	AC117832 Homo sapi
45	44.8	4.7	207804	2	AC011714	AC011714 Homo sapi

ALIGNMENTS

RESULT 1
LOCUS AR077712 954 bp DNA linear PART 31-AUG-2000
DEFINITION Sequence 1 from patent US 5962258.
ACCESSION AR077712
VERSION AR077712.1 GI:10004458
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 954)
AUTHORS Mathur,E.J. and Lam,D.E.
TITLE Carboxymethyl cellulase fromthermotoga maritima
JOURNAL Patent: US 5962258-A 1 05-OCT-1999;
FEATURES Location/Qualifiers

source	1. .954	/organism="unknown"
BASE COUNT	335 a	145 c 221 g 253 t
ORIGIN		
Query Match	100.0%	Score 954; DB 6; Length 954;
Best Local Similarity	100.0%	Pred. No. 3.2e-203;
Matches 954; Conservative	0;	Mismatches 0; Indels 0; Gaps 0;

Oy	1	ATGGGTGTTGATCCCTTTTGAAGAAGACAAAATATTTGGGAAGACGATTAATATATAGAAAT	30
Db	1	ATGGGTGTTGATCCCTTTTGAAGAAGACAAAATATTTGGGAAGACGATTAATATATAGAAAT	60
Oy	61	GGCGTTGAAGACACCAATAGAGGAGACTGGGGAGTGGTATAAAGATGAGTCTTCGAC	120
Db	61	GGCGTTGAAGACACCAATAGAGGAGACTGGGGAGTGGTATAAAGATGAGTCTTCGAC	120
Oy	121	ATTATTAAGAAGACGGGTTCCTCATGTTTCGAATTTCCAAATAGATGAGTACGACGCT	180
Db	121	ATTATTAAGAAGACGGGTTCCTCATGTTTCGAATTTCCAAATAGATGAGTACGACGCT	180
Oy	181	TACGGTTTCCTCTTATTAATAATCATGGAATGCGTCTTCAAGAAGTGAAGTGAAGTAT	240
Db	181	TACGGTTTCCTCTTATTAATAATCATGGAATGCGTCTTCAAGAAGTGAAGTGAAGTAT	240
Oy	241	AACGGAGCCCTTAAGAAGAGGAGACTGGCTTTCCTATAATATTCCTACTACGAGGACTTA	300
Db	241	AACGGAGCCCTTAAGAAGAGGAGACTGGCTTTCCTATAATATTCCTACTACGAGGAGTTA	300
Oy	301	ATGATGATCCAGAAAGACACAAGAAAGATTCTGCTCTTGGAAACAAATTCGCGAT	360
Db	301	ATGAAATGATCCAGAAAGACACAAGAAAGATTCTGCTCTTGGAAACAAATTCGCGAT	360
Oy	361	CGTTATTAAGACTATCCCGAAGACTATTTTTTGAATTTTGAATGAACTCAGCGAAT	420
Db	361	CGTTATTAAGACTATCCCGAAGACTATTTTTTGAATTTTGAATGAACTCAGCGAAT	420
Oy	421	CTTACTCCGGAAAAATGGAATGATGCTTTGAGGAAGCTCTTAAAGTTATTAAGATCAATT	480
Db	421	CTTACTCCGGAAAAATGGAATGATGCTTTGAGGAAGCTCTTAAAGTTATTAAGATCAATT	480
Oy	481	GACAAAAAGACACTTAATTTATAGCAGACAGTGAATGGGGGGTATATTCGCCCTTGA	540
Db	481	GACAAAAAGACACTTAATTTATAGCAGACAGTGAATGGGGGGTATATTCGCCCTTGA	540
Oy	541	AAACTGTCTGTCCCAAAATGGGAAAAAAATCTATAGTTACATTCACACTACAAATCCT	600
Db	541	AAACTGTCTGTCCCAAAATGGGAAAAAAATCTATAGTTACATTCACACTACAAATCCT	600
Oy	601	TTCGAATTTTACCATCAAGAGAGCTGATGGTGGGAAGATCTGACAAATGATGGTGAAGA	660
Db	601	TTCGAATTTTACCATCAAGAGAGCTGATGGTGGGAAGATCTGACAAATGATGGTGAAGA	660
Oy	661	AACTGGGAGTCTCCAGATGATCAGANAACATTTGATAGAGAATTCATTTTATAGAGAA	720
Db	661	AACTGGGAGTCTCCAGATGATCAGANAACATTTGATAGAGAATTCATTTTATAGAGAA	720
Oy	721	TGTCATAAAAAAGACAAGAACCAATTTACATAGAGTATTTGGTGCTCACGAAAACCT	780
Db	721	TGTCATAAAAAAGACAAGAACCAATTTACATAGAGTATTTGGTGCTCACGAAAACCT	780
Oy	781	GACCTTAATACAGAATTAATAANTGAGACTCCTTGTGCTTCGGGAATGAGAAAAAGAGA	840
Db	781	GACCTTAATACAGAATTAATAANTGAGACTCCTTGTGCTTCGGGAATGAGAAAAAGAGA	840
Oy	841	TGAGAGCTGGGCAATCTGGGAATTTTGTCCGGTTTGGTGTATATGATCTCTGAGAAA	900
Db	841	TGAGAGCTGGGCAATCTGGGAATTTTGTCCGGTTTGGTGTATATGATCTCTGAGAAA	900
Oy	901	ACCTGGAATTAAGATCTTTTAGAAGCTTTAATATGAGAGGAGATAGCATTTGAATAA	954
Db	901	ACCTGGAATTAAGATCTTTTAGAAGCTTTAATATGAGAGGAGATAGCATTTGAATAA	954

RESULT 2				
AR096552				
LOCUS	AR096552	954 bp	DNA	linear
DEFINITION	Sequence	1 from patent US 6008032.		
ACCESSION	AR096552			
VERSION	AR096552.1	GI:10025457		
KEYWORDS				
SOURCE	Unknown.			
ORGANISM	Unknown.			
REFERENCE	Unclassified.			
AUTHORS	1 (bases 1 to 954)			
TITLE	Mathur,E.J. and Lam,D.E.			
JOURNAL	Carboxymethyl cellulase from <i>Thermotoga maritima</i>			
FEATURES	Patent: US 6008032-A 1 28-DEC-1999;			
source	location/Qualifiers			
	1..954			

BASE COUNT	335 a	145 c	221 g	253 t	
ORIGIN					
Query Match	100.0%; Score 954; DB 6; Length 954;				
Best Local Similarity	100.0%; Pred. No. 3.2e-203;				
Matches 954; Conservative	0; Mismatches 0; Indels 0; Gaps 0				
QY	1	ATGGGTGTTGATCCTTTTGAAGAGCAAAATATTGGGAAGGCGCATTAATATGAAAT	60		
DB	1	ATGGGTGTTGATCCTTTTGAAGAGCAAAATATTGGGAAGGCGCATTAATATGAAAT	60		
QY	61	GGCGTTGAAGACCAAAATGAGGGAGACTGGGGAGTGGTATAAAGATGAGTTCCTGCAC	120		
DB	61	GGCGTTGAAGACCAAAATGAGGGAGACTGGGGAGTGGTATAAAGATGAGTTCCTGCAC	120		
QY	121	ATTATTAAGAAGAGCGGCTTCCTCATGTTGCAATTCGAATTCGAATGAAGATGAGATGCCAGGCT	180		
DB	121	ATTATTAAGAAGAGCGGCTTCCTCATGTTGCAATTCGAATGAAGATGAGATGCCAGGCT	180		
QY	181	TACGGCTTTCCTCTTATAAATCATGAGATGCGCTCTCAAAAAGAGTGAATGAAGTATA	240		
DB	181	TACGGCTTTCCTCTTATAAATCATGAGATGCGCTCTCAAAAAGAGTGAATGAAGTATA	240		
QY	241	AACGGAGCCGTGAAGAAGAGAGACTGGCTGTGCTATATAATATTCATCAGACAGAGTTA	300		
DB	241	AACGGAGCCGTGAAGAAGAGAGACTGGCTGTGCTATATAATATTCATCAGACAGAGTTA	300		
QY	301	ATGAATGATCCAGAGAAGACACAGAAGAAATTTCTTGCTCTTTGGAAACAATTGCTGAT	360		
DB	301	ATGAATGATCCAGAGAAGACACAGAAGAAATTTCTTGCTCTTTGGAAACAATTGCTGAT	360		
QY	361	CGTTATTAAGACTATCCCGAAACTCTATTTTTGAATTCGTGAATGCAACCTCAGCGAAT	420		
DB	361	CGTTATTAAGACTATCCCGAAACTCTATTTTTGAATTCGTGAATGCAACCTCAGCGAAT	420		
QY	421	CTTACCCGGGAAAAATGGAATGAAGCTGTTGAGAAAGCTCTAAAGTTTAAAGATCAATT	480		
DB	421	CTTACCCGGGAAAAATGGAATGAAGCTGTTGAGAAAGCTCTAAAGTTTAAAGATCAATT	480		
QY	481	GACAAAAACACACATATATTAATTAAGCACACAGCTGAATGGGGGGTATATCTGCCCTTGA	540		
DB	481	GACAAAAACACACATATTAATTAAGCACACAGCTGAATGGGGGGTATATCTGCCCTTGA	540		
QY	541	AAACTGTCTGCCCAAAATGGGAAAAAAATCTTACCTTCAATTACATCTACATCAATCCT	600		
DB	541	AAACTGTCTGCCCAAAATGGGAAAAAAATCTTACCTTCAATTACATCTACATCAATCCT	600		
QY	601	TTTGCAATTTACCATCAAGAGACTAGTGGGTGAAGAGACTCAGAAATGGTTGGGAGAA	660		
DB	601	TTTGCAATTTACCATCAAGAGACTAGTGGGTGAAGAGACTCAGAAATGGTTGGGAGAA	660		
QY	661	AAGTGGGAGATCTCCAGATGATCAGAAACATTTGTATAGAGATTCATTTTATAGAAGAA	720		
DB	661	AAGTGGGAGATCTCCAGATGATCAGAAACATTTGTATAGAGATTCATTTTATAGAAGAA	720		

QY 721 TGGTCAAAAAGAACAAACCAATTTACATAGTGTAGTTGGTCCCTACAGAAAAAGCT 780
 |||
 Db 721 TGGTCAAAAAGAACAAACCAATTTACATAGTGTAGTTGGTCCCTACAGAAAAAGCT 780
 QY 781 GACCTTGATCAAGATTAATAATGAGCTCTTGTGTGTGCGCAATTCGAGAAAAGAGAGA 840
 |||
 Db 781 GACCTTGATCAAGATTAATAATGAGCTCTTGTGTGTGCGCAATTCGAGAAAAGAGAGA 840
 QY 841 TGGAGCTGGGCATCTGAGCAATTTGTCGGCTTTGGTGTATGATCTGAGAAAA 900
 |||
 Db 841 TGGAGCTGGGCATCTGAGCAATTTGTCGGCTTTGGTGTATGATCTGAGAAAA 900
 QY 901 ACCTGGAATTAAGATCTTTAGAGCTTTAATAGAGAGAGATGACATTTGAATA 954
 |||
 Db 901 ACCTGGAATTAAGATCTTTAGAGCTTTAATAGAGAGAGATGACATTTGAATA 954

RESULT 3
 ARI57661
 LOCUS ARI57661 954 bp DNA linear PART 17-OCT-2001
 DEFINITION Sequence 1 from patent US 6245547.
 ACCESSION ARI57661
 VERSION ARI57661.1 GI:16218624
 KEYWORDS
 SOURCE Unknown.
 ORGANISM Unknown.
 REFERENCE 1 (bases 1 to 954)
 AUTHORS Mathur,E.J. and Lam,D.E.
 TITLE Carboxymethyl cellulose from thermotoga maritima
 JOURNAL Patent: US 6245547-A 1 12-JUN-2001;
 FEATURES
 source 1..954
 /organism="unknown"

BASE COUNT 335 a 145 c 221 g 253 t
 ORIGIN
 Query Match 100.0%; Score 954; DB 6; Length 954;
 Best Local Similarity 100.0%; Pred. No. 3.2e-203;
 Matches 954; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGGGCTGTGATCTTTTGAAGAACAAATTTGGAGAGGACATTAATATAGAAAT 60
 |||
 Db 1 ATGGGCTGTGATCTTTTGAAGAACAAATTTGGAGAGGACATTAATATAGAAAT 60
 QY 61 GCGCTTGAGCAACCAATGAGGAGCTGGGAGTGTGATTAAGATAGTCTTCGAC 120
 |||
 Db 61 GCGCTTGAGCAACCAATGAGGAGCTGGGAGTGTGATTAAGATAGTCTTCGAC 120
 QY 121 ATTATTAAGAAAGCCGGTTCTCTCATGTTGCAATTCGAATAGATGAGTACGACGCT 180
 |||
 Db 121 ATTATTAAGAAAGCCGGTTCTCTCATGTTGCAATTCGAATAGATGAGTACGACGCT 180
 QY 181 TACGCTTTCCTCTTATTAATATCATGATCGCTTCTTCAAAAAGAGTGAAGTGATA 240
 |||
 Db 181 TACGCTTTCCTCTTATTAATATCATGATCGCTTCTTCAAAAAGAGTGAAGTGATA 240
 QY 241 AACGAGCCCTGTAAGAGAGAGCTGGCTGTGATTAATATTCATACAGAGAGTTA 300
 |||
 Db 241 AACGAGCCCTGTAAGAGAGAGCTGGCTGTGATTAATATTCATACAGAGAGTTA 300
 QY 301 ATGAATGATCCAGAAAGACACAAAGAAAGATTTCTGCTCTTTGGAACAAATTCGTGAT 360
 |||
 Db 301 ATGAATGATCCAGAAAGACACAAAGAAAGATTTCTGCTCTTTGGAACAAATTCGTGAT 360
 QY 361 CGTTATAAAGACTATCCGGAAGTCTATTTTGAATTTCTGATAGTACCTCAGGAAT 420
 |||
 Db 361 CGTTATAAAGACTATCCGGAAGTCTATTTTGAATTTCTGATAGTACCTCAGGAAT 420
 QY 421 CTTACTCCGGAATAATGATGATGACTGTGAGGAGCTCTAAGATTAATAGATCAAT 480
 |||
 Db 421 CTTACTCCGGAATAATGATGATGACTGTGAGGAGCTCTAAGATTAATAGATCAAT 480

QY 481 GACAAAAAGCACATTAATTTATAGGCACAGCTGAATGGGGGGTATATTCGCCCTTGA 540
 |||
 Db 481 GACAAAAAGCACATTAATTTATAGGCACAGCTGAATGGGGGGTATATTCGCCCTTGA 540
 QY 541 AAACGTCTCTCCCAAAATGGGAAAAAATTCTATAGTCAATTCATCTACATCCT 600
 |||
 Db 541 AAACGTCTCTCCCAAAATGGGAAAAAATTCTATAGTCAATTCATCTACATCCT 600
 QY 601 TTGCAATTTCCCATCAAGAGAGCTGGTGGTGAAGGATCTGAGAAATGGTGGGAGA 660
 |||
 Db 601 TTGCAATTTCCCATCAAGAGAGCTGGTGGTGAAGGATCTGAGAAATGGTGGGAGA 660
 QY 661 AACTGGGATCTCCAGATGATGATCAAGAACATTTGATAGAAATTCATTTATAGAA 720
 |||
 Db 661 AACTGGGATCTCCAGATGATGATCAAGAACATTTGATAGAAATTCATTTATAGAA 720
 QY 721 TGGTCAAAAAGAACAAACCAATTTACATAGTGTAGTTGGTCCCTACAGAAAAGCT 780
 |||
 Db 721 TGGTCAAAAAGAACAAACCAATTTACATAGTGTAGTTGGTCCCTACAGAAAAGCT 780
 QY 781 GACCTTGATCAAGATTAATAATGAGCTCTTGTGTGTGCGCAATTCGAGAAAAGAGAGA 840
 |||
 Db 781 GACCTTGATCAAGATTAATAATGAGCTCTTGTGTGTGCGCAATTCGAGAAAAGAGAGA 840
 QY 841 TGGAGCTGGGCATCTGAGCAATTTGTCGGCTTTGGTGTATGATCTGAGAAAA 900
 |||
 Db 841 TGGAGCTGGGCATCTGAGCAATTTGTCGGCTTTGGTGTATGATCTGAGAAAA 900
 QY 901 ACCTGGAATTAAGATCTTTAGAGCTTTAATAGAGAGAGATGACATTTGAATA 954
 |||
 Db 901 ACCTGGAATTAAGATCTTTAGAGCTTTAATAGAGAGAGATGACATTTGAATA 954

RESULT 4
 AE001813
 LOCUS AE001813 10244 bp DNA linear BCT 02-JUN-1999
 DEFINITION Thermotoga maritima section 125 of 136 of the complete genome.
 ACCESSION AE001813 AE000512
 VERSION AE001813.1 GI:4982321
 KEYWORDS
 SOURCE Thermotoga maritima.
 ORGANISM Thermotoga maritima.
 REFERENCE 1 (bases 1 to 10244)
 AUTHORS Nelson K.E., Clayton,R.A., Gill,S.R., Gwyn,M.L., Dodson,R.J., Haft,D.H., Hickey,E.K., Peterson,J.D., Nelson,W.C., Ketchum,K.A., McDonald,L., Utterback,T.R., Malek,J.A., Linher,K.D., Garrett,M.M., Stewart,A.M., Cotton,M.D., Pratt,M.S., Phillips,C.A., Richardson,D., Heidelberg,J., Sutton,G.G., Fleischmann,R.D., Eisen,J.A., Fraser,C.M. et al.
 TITLE Evidence for lateral gene transfer between Archaea and bacteria from genome sequence of Thermotoga maritima
 JOURNAL Nature 399 (6734), 323-329 (1999)
 MEDLINE 99287316
 PUBMED 10360571
 REFERENCE 2 (bases 1 to 10244)
 AUTHORS Nelson K.E., Clayton,R.A., Gill,S.R., Gwyn,M.L., Dodson,R.J., Haft,D.H., Hickey,E.K., Peterson,J.D., Nelson,W.C., Ketchum,K.A., McDonald,L., Utterback,T.R., Malek,J.A., Linher,K.D., Garrett,M.M., Stewart,A.M., Cotton,M.D., Pratt,M.S., Phillips,C.A., Richardson,D., Heidelberg,J., Sutton,G.G., Fleischmann,R.D., White,O., Salzberg,S.L., Smith,H.O., Venter,J.C. and Fraser,C.M.
 TITLE Submitted (01-JUN-1999) The Institute for Genomic Research, 9712 Medical Center Dr. Rockville, MD 20850, USA
 JOURNAL Location/Qualifiers
 FEATURES
 source 1..10244
 /organism="Thermotoga maritima"
 gene /db_xref="taxon:2336"
 73..1104
 /gene="TM1744"
 73..1104
 CDS

gene
CDS
/gene="TM1744"
/note="similar to GB:U00096 SP:P75823 PID:1787095
PID:2225892 percent identity: 67.58; identified by
sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="L-allo-threonine aldolase"
/protein_id="AAD36809.1"
/db_xref="GI:4982322"
/translation="MIDLRSPVTAKPTEEMKKAQAEVGDVYGEDPTINELRLAA
ETREKALVPVSTMGNOVSIMAHTRGDVITLADSHIFWEVGAAVAGVMPHP
VPGNGAMDDVDVKARIPRNIHPRPSLIAIEMTHRSRGRVPLEMKICITAKE
HGINVHIDGARIFENASIASGVPVEYAGIADSVFCISKGCAPAGSVYVDDRIER
AKARKMLGGMRQAGVLAAGIATLKMVDLKEDEHNAFLKLKEIGSYNPE
VKTMMILRTDNLKVNAGFIEALRNSGVLANAVSDEIRLVTHKVSRIEELANI
FEKLFKES"

gene
CDS
1183..1806
/gene="TM1745"
1183..1806
/gene="TM1745"
/note="similar to percent identity: 0.00; identified by
sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAD36810.1"
/db_xref="GI:4982323"
/translation="MKRFVVMVLVTVALGVFAAGYRGFAERAPEROPMGYAOQFG
CGGGIGVGNAPMYERQPMARERGVGKIQIRDEQLKALKRIEALNNIDLD
DOLKEIVNAKETKEQLDLKEKEFQLHEEYVNALVKRDTKAEIDLKQIDLRDEVO
DIKDFAKVDIITVDYKRFGRGKRLFYLLIDEGEVELEEMQ"

gene
CDS
2074..4002
/gene="TM1746"
2074..4002
/gene="TM1746"
/note="similar to GB:AE000783 percent identity: 56.54;
identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="oligopeptide ABC transporter, periplasmic
oligopeptide-binding protein"
/protein_id="AAD36811.1"
/db_xref="GI:4982324"
/translation="MRKFVGFVLVLAFAFYSEVEVQLPMFVNTFDPYVQAQHL
AGTFFEGVGPDPKSLGVKVVPAVCESTYSEDDLTTFYLRKDRMNGEPVT
AYDEFPAKRASPTPTPLPMASPVQYIKYVYDCKSGAKPLDELGVKRLDYTLOVT
LSRPSTINMLVGLGAMPILKRTYOEHPEDMKRPYVPGNCPYIESTTPYVEITLV
RNKYVGVDPGNVDRIVLAKAGLQOYLAMGIDAVFTVAGDIYFALKNKQLSKELH
ESGVQWVGYEITRSLSPVFDLIRIKALAMIDKVLVDIVLGMAIPTHAYCSPDS
ETAEAVNGIYPDEQAKLLAEAGYPNGKFPKVFYITGASDPVAEALVDQMKVVG
ITFEENIESGGSTLQMSDFVMAEPGCTTGGPVNMQDSLSMQTADHTWFDYF
AEVRGKIYELKRESYLREEGKTDEEKLEILNRAQELMEKMOEYVNEPDEFLKX
FLPAPGELTFTEQINKLYEEMKATTDOKQLWRLAMRAYVNOEYSLVEYTGMPY
NNEARLWARILYSPDKAVEYAKKLQQLCVDLATEVPLVQKVIYLAFLYKGIYF
KESMGTFELFYLYVEK"

gene
CDS
3999..5024
/gene="TM1747"
3999..5024
/gene="TM1747"
/note="similar to GB:I42023 SP:PA5054 PID:1006439
PID:1221243 PID:1205369 percent identity: 65.90;
identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="oligopeptide ABC transporter, permease protein"
/protein_id="AAD36812.1"
/db_xref="GI:4982325"
/translation="MNSPEWDPGLLNNNSGVDMRFLKRLTLTAISNVVIV
ITYVLMWLAGNFEELQVRYDAIARTPDDPAYQATLKGFEERYGLNPLMQILMY
LKGAVMKRGPSPDPARNIEDLKEKPTPTFLASLILFALVGVPLGILAAKLN
TWIDYRAMVSVYGVAPISVYVAVPLILFISYILGMPLTSGMEGIRKILPTIALAKN
PLASVARTSVSLDTLNDQFIRTAFAKSGDDTVTGMKHALRSMPIYITVPGMAV
LWGVYVENIFRIPGLGLFANAAYTRDIPLVITSTFILATVIMNMLIVDLVAIL

gene
CDS
5042..5911
/gene="TM1748"
5042..5911
/gene="TM1748"
/note="similar to GB:M57689 SP:P24139 GB:X56347 PID:143606
PID:40007 percent identity: 64.69; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/product="oligopeptide ABC transporter, permease protein"
/protein_id="AAD36813.1"
/db_xref="GI:4982326"
/translation="MLKKNKTESGYWKAFLRKKKNAVIGVFLILALAILAP
YIAPYDPDEHYIRAPGSPKDFIRFGDALGRPLFSRIIYSRNCIIGEGSOPVLI
IGGLGVAGPKGKSIDKFTMSIVDIMEAFPPLEWITVTLWAGRLGFLPTLAGITG
WAGMARLVQGVLYLKNSEFEVAKAGASTFTIRKHLIPNMGILVNLAFVPGA
MTESGLAIVIGMGVRRPPSMGNLIGGIGMMAFPFLIFPAVTFATLISFLAD
GLRDAFNRSEIS"
5908..6882
/gene="TM1749"
5908..6882
/gene="TM1749"
/note="similar to SP:P04285 GB:X05491 GB:X52093 PID:47805
percent identity: 69.65; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/product="oligopeptide ABC transporter, ATP-binding
protein"
/protein_id="AAD36814.1"
/db_xref="GI:4982327"
/translation="MMELNVMNLKVEFHRVEGIVKAVDGISYKLNKGSLGIVGSG
SKGSVSIILRLINRNGRIYDGAFLIGDGLKLNKEELRIGRIDISIFQNPMTS
LNPILIRGIVMEPIIWRMLKNEAEERAIETELLERGIPESPKRLNFPQSGMAR
QRVIMALACHKRLIADDEFTTALDVTIOAQIMEILOETKEEYGSVFIFTHDSLVA
TNFDRITITWAGKIVEAEVEELKTLPHFTYTGILNSTLEIGSGKRLVPIPGNP
NPKRPSGCKPHPCSFAMEICQREPPLVNISNNKVAACHLIGESK"
6852..7838
/gene="TM1750"
6852..7838
/gene="TM1750"
/note="similar to SP:PA2065 PID:677944 GB:AL009126 percent
identity: 76.43; identified by sequence similarity;
putative"
/codon_start=1
/transl_table=11
/product="oligopeptide ABC transporter, ATP-binding
protein"
/protein_id="AAD36815.1"
/db_xref="GI:4982328"
/translation="MPEDQGIKKKPLQIYDLKKYPPQCKRIKANDGISIEIKKEE
TLGLVSGCGSKSLGRILKILRLPDGKIFEEKDLTNLDMKMYRKMKQIIFOD
PLGSINPOMYGRILIEDPLLIHKIGTKERRRVEELDWIGIGREFINSFPHEFSG
QOORIGIALALANLKPFLVCEPVSALDVSIOAQIIDLEPIQCKMSISVLFIAHNL
VYEHSHKVAWYVICKIYEVGDVQKIFLNPHTRALKVKPKIPMDGOKORFYSLK
GLDESDIDLPKRCRQTCTCKKALCFKEPELLEVENHVSCHLVRSRG"
7898..8851
/gene="TM1751"
7898..8851
/gene="TM1751"
/note="similar to SP:P23340 percent identity: 52.35;
identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="endoglucanase"
/protein_id="AAD36816.1"
/db_xref="GI:4982329"
/translation="AGVDPFENKLTIGGINIGNALPAPNEGDMGVYIKDFEPIIKE
AFSHVRIIRNSTHAIYAPPKINDREFKRYDVTYNALRGAAVYNIHHYELMN
DDEHKERFLAMKQIADYKYDPELTLEILNEHGLNTEPKNELEELKLYIRST
DKKHIIIGTAWGGISALEKISVPRKENSIVTTHNPEFTHOGAEVGGSEKWL
GRKMGSPDQKHLIEEFNIEBWSKKNRPYIYIGFAYRRADLESRIKTSFVREM

gene
EKRRSMAYWFCGSGFYVDTLRKTNKDLLEALIGDSIE"
8844..9833
/gene="TM1752"
8844..9833
/gene="TM1752"
/note="similar to SP:P23340 percent identity: 51.51;
identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="endoglucanase"
/protein_id="AAD36817.1"
/db_xref="GI:4982330"
/translation="NNNTTPRWGRFNLLEAFSISKTGNFKEEDFLMMAQWENFVRIP
MCHLMSDRGNPFIIREDPEKIDRIVYGEYGIHICISLRAGSYNKEVEKTN
LWKDETAEAFIHMNSFIARRYKGISSTHLENPEPDPQIMSVEDHNSLIKRT
ITEIRKIDPERLIIDIGYGNIPVDLIENTVOSCHGYIPFSVTHVKAWEVDSKDF

Query Match 99.8%; Score 952.4; DB 1; Length 10244;
Best Local Similarity 99.9%; Pred. No. 5,5e-203;
Matches 953; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

1 ATGGGTGTTGATCCTTTGAAAGACAAATTTGGGAAGAGCATTAATATAGGAAT 60
|||||
7898 ATGGGTGTTGATCCTTTGAAAGACAAATTTGGGAAGAGCATTAATATAGGAAT 7957
|||||
61 GCGCTTGAAGACCAATGAGGAGAGTGGGAGTGTATTAAGATGATGTTCTTCGAC 120
|||||
7958 GCGCTTGAAGACCAATGAGGAGAGTGGGAGTGTATTAAGATGATGTTCTTCGAC 8017
|||||
121 ATTATTAAGAACCGCGTTCTCTCATGTTTCGATTTCCAAATAAGATGATGACGACGCT 180
|||||
8018 ATTATTAAGAACCGCGTTCTCTCATGTTTCGATTTCCAAATAAGATGATGACGACGCT 8077
|||||
181 TAGCGTTCTCTCATGTTTCGATTTCCAAATAAGATGATGACGACGCT 240
|||||
8078 TAGCGTTCTCTCATGTTTCGATTTCCAAATAAGATGATGACGACGCT 8137
|||||
241 AACGAGCCCTGAAAGAGAGTGGTGTGCTATTAATATTCATCACTACGAGAGATTA 300
|||||
8138 AACGAGCCCTGAAAGAGAGTGGTGTGCTATTAATATTCATCACTACGAGAGATTA 8197
|||||
301 ATGAATGATCCAGAGACACACAGAGAAATTTCTGCTCTTTGGAAACAATTTGCTGAT 360
|||||
8198 ATGAATGATCCAGAGACACACAGAGAAATTTCTGCTCTTTGGAAACAATTTGCTGAT 8257
|||||
361 CATTATTAAGACATCCGAAACTCTATTTTGAATTTCTGAATGAACCTCAGCGAAT 420
|||||
8258 CATTATTAAGACATCCGAAACTCTATTTTGAATTTCTGAATGAACCTCAGCGAAT 8317
|||||
421 CTTACTCCGAAAAAATGGAATGAATCTTGAAGAGCTCTAAAGTTATTAAGATCAAT 480
|||||
8318 CTTACTCCGAAAAAATGGAATGAATCTTGAAGAGCTCTAAAGTTATTAAGATCAAT 8377
|||||
481 GACAAAAACACACTTAATTTATAGCACAGCTGAATGGGGGGGTATATCTGCCCTTGA 540
|||||
8378 GACAAAAACACACTTAATTTATAGCACAGCTGAATGGGGGGGTATATCTGCCCTTGA 8437
|||||
541 AACACTCTGTCCCAAAATGGGAAAAAATTCATAGTTACAAATTCAGACTACATCCG 600
|||||
8438 AACACTCTGTCCCAAAATGGGAAAAAATTCATAGTTACAAATTCAGACTACATCCG 8497
|||||
601 TTGGAATTTACCATCAAGAGAGTGGTGGGAGATCTGAGAAAGGTTGGGAGAA 660
|||||
8498 TTGGAATTTACCATCAAGAGAGTGGTGGGAGATCTGAGAAAGGTTGGGAGAA 8557
|||||
661 AAGTGGGATCTCCAGATGATCAGAAACATTTGATAGGAATTCATTTATAGAGAA 720
|||||
8558 AAGTGGGATCTCCAGATGATCAGAAACATTTGATAGGAATTCATTTATAGAGAA 8617
|||||
721 TGGTCAAAAAAGACCAATTTTACATAGGTGAGTTGGTCCCTACGAAAAAGCT 780
|||||
8618 TGGTCAAAAAAGACCAATTTTACATAGGTGAGTTGGTCCCTACGAAAAAGCT 8677

QY 781 GACCTTGATCAAGATTAATAATGACCTCCTTGTGTCGCGAAATGAGAAAGAGAG 840
|||||
DB 8678 GACCTTGATCAAGATTAATAATGACCTCCTTGTGTCGCGAAATGAGAAAGAGAG 8737
|||||
QY 841 TGGAGCTGGGATACGCGAATTTTGTCCGGTGTGGTGTATATATCTGAGAGAA 900
|||||
DB 8738 TGGAGCTGGGATACGCGAATTTTGTCCGGTGTGGTGTATATATCTGAGAGAA 8797
|||||
QY 901 ACTGGAATTAAGATCTTTAGAGCTTTAATAGAGAGATAGCATTTGATTA 954
|||||
DB 8798 ACTGGAATTAAGATCTTTAGAGCTTTAATAGAGAGATAGCATTTGATTA 8851
|||||

RESULT 5
AR205090 1083 bp DNA linear PAT 20-JUN-2002
LOCUS
DEFINITION Sequence 13 from patent US 6368844.
ACCESSION AR205090
VERSION AR205090.1 GI:21502584
KEYWORDS
SOURCE
ORGANISM Unknown.
REFERENCE Unclassified.
1 (bases 1 to 1083)
AUTHORS Bylina, E.J.
TITLE Glycosidase enzymes
JOURNAL Patent: US 6368844-A 13 09-APR-2002;
FEATURES Location/Qualifiers
source
1..1083
/organism="unknown"
BASE COUNT 367 a 189 c 242 g 285 t
ORIGIN

Query Match 24.7%; Score 235.8; DB 6; Length 1083;
Best Local Similarity 54.2%; Pred. No. 1.4e-42;
Matches 502; Conservative 0; Mismatches 422; Indels 3; Gaps 1;

14 CTTTGAAGAGAACAAATATTGGAAGAGCATTAATATAGAGAAATGCCCTTGAAGCAC 73
|||||
149 CATTGAATTAACAAATATGCTAGTAAGAGATTAATATGGAATGCTTTGAAAGCTC 208
|||||
QY 74 CAATGAGGAGACCTGGGAGTGTGATTAAGATGATGATGCTTCGACATTAATAAGAG 133
|||||
DB 209 CTTTGAAGAGACCTGGGAGTGTGATTAAGATGATGATGCTTCGACATTAATAAGAG 268
|||||
QY 134 CCGGTTCTCTCATGTTTCGAATTCATTAAGATGAGAGTACGACGCTTACGCGTTCTC 193
|||||
DB 269 GGGGATTGATCTGTAGGATTCCTATTAAGATGATGATGATGATGATGATGATGATG 328
|||||
QY 194 CTTATTAATTCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 253
|||||
DB 329 CATATATATTTGACAGAAATTTCTCGAAGAGTAAACATGTTGTGATAGAGGCTCTTG 388
|||||
QY 254 AAAGAGAGCTGGTGTCTGATTAATATTCATCACTACGAGAGATTAAGATGATGATG 313
|||||
DB 389 AGAATTAATTAACAGTAATCATCAATACGACCATTTTGAAGACCTCTATCAAGAACCG 448
|||||
QY 314 AAGAACACAAAGAAATTTCTGCTTTGGAACAAATTTGCTGATTCGTTATAAGACT 373
|||||
DB 449 ATTAATACGCGGATGTTTGGTGAATTTGGAGACAGATTTGAAATTTCTTTAAAGATT 508
|||||
QY 374 ATCCCAAACTCATTTTGTGAAATTTCTGAATGACCTCAAGGAATTTCTACTCCGAGAA 433
|||||
DB 509 ACCCGGAATATCTGTTCTTTGAAATTTCTACAGAGAGCTCTCGAAGCTTGACAGCTGAAA 568
|||||
QY 434 AATGGAATGACTGCTTGAAGAGCTCTAAAGATTAAAGATCAATTTGAACAAAGACACA 493
|||||
DB 569 AATGGAATGACTGCTTGAAGAGCTCTAAAGATTAAAGATCAATTTGAACAAAGACACA 628
|||||
QY 494 CATATATTTAGGACAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATG 553
|||||
DB 629 TTGCTATTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 688
|||||

QY 554 CAAATGGAAGAAAAATCTATAGTCAATTCACATCCATCCCTTGGATTTACCC 613
 11
 Db 689 TCAA---CGACAAACGATCATCTCTTCCTTCCATCTACAGACCTTCAATTCACAC 745
 QY 614 ATCAAGAGCTAGCTGGGTGGAAGAGATCTGAGAAATGTTGGGAGAAAGAGGATCTC 673
 11
 Db 746 ATCAGGCTGCCGAATGGGTAAATCCATCCACCTGTTAGGTTAAGTGGATGGCGAGG 805
 QY 674 CAGATGATCAAGAACATTTGATAGAGAAATCAATTTTATAGAAAGATGTCAGAAAAA 733
 11
 Db 806 AATGGGAATTAACCAATACGAAGTCAATTCATTAACGTAGAGACGGGCAAGCAAA 865
 QY 734 ACAAGAGCAATTTACATAGATGATTTGGTCCCTACAGAAAGCTACCTGATTAAT 793
 11
 Db 866 ATACGATACCAATCTCTTCTGTAATCGCTGTTATCAAAAGCAGACATGAGCTCA 925
 QY 794 GAATAAATGACCTCTTGTCTGTCGGAATGAGAAAGAGATGAGATGAGCTGCGAT 853
 11
 Db 926 GGGTAAAGTGAACCGAAGATGTGAGAAATGGCGGAAGATTTGGATTTTCATCGCGT 985
 QY 854 ACTGGGAATTTTGTCCGGTTTGTGTTATGATATCTGAGAAAGCTGGAATTAAG 913
 11
 Db 986 ATGGGAATTTTGTGACAGATTTGGCATATAGATAGATGTCTCAAACTGGATGCAAC 1045
 QY 914 ATCTTTAGAGCTTTATAGAGAG 940
 11
 Db 1046 CATTGGCAACAGCTGTGTTGGCACAG 1072

RESULT 6

CLOCCELCD 2160 bp DNA linear BCT 01-FEB-2000
 LOCUS C.cellulolyticum celCCD gene for endo-1,4-glucanase (EC 3.2.1.4),
 DEFINITION complete cds.

ACCESSION D90341
 VERSION D90341.1 GI:216414
 KEYWORDS celCCD gene; endo-1,4-glucanase.
 SOURCE C.cellulolyticum (ATCC 35319) DNA, clone pAS1.
 ORGANISM Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
 Clostridium.

REFERENCE 1 (bases 1 to 2160)
 AUTHORS Shima,S., Igarashi,Y. and Kodama,T.
 TITLE Nucleotide sequence analysis of the endoglucanase-encoding gene,
 celCCD, of Clostridium cellulolyticum
 JOURNAL Gene 104 (1), 33-38 (1991)
 MEDLINE 92009193

COMMENT These data kindly submitted in computer readable form by: Seigo

Shima
 Department of Biology, Abiko Research Laboratory
 Central Research Institute of Electric Power Industry 1646 Abiko,
 Abiko-city
 Chiba 270-11
 Japan
 Phone: 0471-82-1181
 Fax: 0471-82-7922.

FEATURES
 source
 1. 2160
 /organism="Clostridium cellulolyticum"
 /db_xref="taxon:1521"

-35_signal
 -10_signal
 RBS
 CDS
 /note="SD sequence"
 281..2035
 /note="endo-1,4-glucanase precursor"
 /codon_start=1
 /transl_table=11
 /protein_id="BA14354.1"
 /gb_xref="GI:216415"
 /translation="MKKILAIISCSIIISFLPMYGAINSODVKKMGIGMILCNT
 FPAITRSGSKAOGIEYFDPDFKAGFKHYRPIRMQHLTANSPPYVDSFLRIETV
 IOMLSRGFTVYINSHDITWMDNYSQNGRPFKIMEQIAQRKGSNLYFELNEP

HGNITDSOINDMKRIILNIRKTNPTNVIIGAGWMSYNSLSOLEIPNDPLIATFH
 YXDRPSFTHQMGWTGCTKNDMDALAMVNEHKMSDKNNIPVLYGEEVGWGHSDRTSA
 VMDFVSDQAIASHFSGGANDNGVSGVNDMDAFYNPDTPQFPEKLNLITGTTY
 DMTPPEINDDPKPTPATPATGAEQILIEPEBAMQMAAASVDATASKISSGNNGL
 EITVAGSSNGWGVNDNEHRNDMEKWKIISFDIKSSINFEVRLILAEQKISBDE
 HMTYVLRPSTSWTIIIEIPSSFTKRMDOQPAODGSEFTDYLRVGSJLHPMSNSGT
 LINDIKILGPEEIGIKIDVNEGDNIDAIDRALLKRYLLDSSISINVMNDIND
 GDMNIDFAKLKMLGDP"
 281..352
 /note="endo-1,4-glucanase signal peptide"
 353..2032
 /product="endo-1,4-glucanase mature peptide"
 BASE COUNT 765 a 317 c 417 g 661 t
 ORIGIN

Query Match 16.1%; Score 153.2; DB 1; Length 2160;
 Best Local Similarity 54.28; Pred. No. 3.9e-24;
 Matches 333; Conservative 0; Mismatches 278; Indels 3; Gaps 1;

QY 4 GGTGTGATCCTTTTGAAGGAACAAATATTTGGAGAGGCAATTAATAGGAATGCG 63
 11
 Db 353 GCTATCAATCTCGAGATATGTAATAAAGATGGAATGAACTTGGAAATACG 412
 QY 64 CTGGAACACCAATGAGGAGACTGGGAGCTGTGATAAAGATGATGTTCTGCACAT 123
 11
 Db 413 TTTCAGCACCACAGGAAGGTTCTTGGTCAAGGACGCTCAGAGTACTATTTGATGAT 472
 QY 124 ATAAAGAGAGCGGTTTCTCATGTTCGAATTCATTAAGATGAGTACGACGCTTAC 183
 11
 Db 473 TTAAACAGCGCGGTTTCAAGCATGTAGAAATCCGTTGGGATCAGCATACATT 532
 QY 184 GCGTTTCTCTTATTAATATCATGATCGTCTTCAAAAGATGATGATGAGTAAAC 243
 11
 Db 533 GCAACAGTCGCTATGCTGTGACAGCAACTCTTAAACAGAAATGAACTGTAATAGAC 592
 QY 244 GAGGCCCTGAAAGAGGAGCTGGTGTGCTATTAATATCATCATACGAGGATTAAAG 303
 11
 Db 593 TGGTACTTCTCCGCTGAGATTGTGACGTAATTAATTCATCATCATGATGATGCTTAA 652
 QY 304 AATGATCCAGAAACACACAGGAAGATTTCTGCTCTTGGAAACAAATGCTGATCGT 363
 11
 Db 653 GATTAATATGACAGAAATATCGAGCGGTTGAAAATAATGGAACAAATGACACACGT 712
 QY 364 TATAAGACTATCCCGGAACCTATTTTGAATTCGATGAACCTCAGGAATCTT 423
 11
 Db 713 TTTAAAGCAAGCTCGAAATCTGGTTTGTGAATTAATTAATGACCAATGTAATATA 772
 QY 424 ACTCGGAAAAATGGAATGAACTGCTTGAGGAGAGCTTAAAGTTATAGATCAATTGAC 483
 11
 Db 773 ACAGATAGCCAGATATAGATATGAAACAAAGAAATTTAATTAATTAAGAAAGACTAAT 832
 QY 484 AAAAGCACATATATATATATAGCAGACGTGAATGGGGGTATATCTGCCCTGAAAA 543
 11
 Db 833 CCGACTAGAAATGTAATTAATTAATGAGGAGTACGGAATACCTATTAATCTTAAGTACG 892
 QY 544 CTGTCTGCCCAAAATGGAATAATTTCTATAGTCAATCTACTACATCAATCTTTC 603
 11
 Db 893 TTGGAATTTCCAAA---CGACCAATATCTATGCAACTTTTCACTATAGACCATAC 949
 QY 604 GAATTAACCATCA 617
 11
 Db 950 TCTTTCACTCACCA 963

RESULT 7

CLOCCELH 3002 bp DNA linear BCT 26-APR-1993
 LOCUS C.Thermocellum cellulase (celH) protein genes, complete cds.
 DEFINITION M31903
 ACCESSION M31903.1 GI:144773
 VERSION M31903.1
 KEYWORDS cellulase; endoglucanase.
 SOURCE C.thermocellum (strain NCIB 10682) DNA.
 ORGANISM Clostridium thermocellum

Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;

Clostridium.

REFERENCE 1 (bases 1 to 3002)

AUTHORS Yague,E., Begun,P. and Aubert,J.P.

TITLE Nucleotide sequence and deletion analysis of the cellulase-encoding

gene celH of Clostridium thermocellum

JOURNAL Gene 89 (1), 61-67 (1990)

MEDLINE 90323606

PUBMED 2197182

COMMENT Draft entry and computer-readable sequence for [1] kindly submitted

by P. Begun, 07-FEB-1990.

FEATURES

Location/Qualifiers

1..3002

/organism="Clostridium thermocellum"

/db_xref="taxon:1515"

159..164

/note="ribosome binding site"

171..2873

/note="cellulase precursor (EC 3.2.1.4)"

/codon_start=1

/transl_table=11

/protein_id="AA023225.1"

/db_xref="GI:144774"

/translation="MKRLVSEFLVSLITVGLLSQSLGNTNSGLKIGAMGTPSES

AIKSFQELQKRLDIVHOFINMSTDSWVRPADAVYNNGLSMITMWEWETVDIK

NGKADAYIRMAQDKKAYKEIIRPLRHANGDMVPAALNGSIRMTNETTYAARHI

VDIFRANGATNVKWNVCNCDVNGSTSLGHPGDNYVDYSIDGYNMGTSQSGSO

WQSPFOVSRATQALASIKPILIIAEPASAEIGKARITTAHSIRSYKIVAAV

WHENKEDIMRINSSPEALAAIREAIGGSSPPTPTPTWSTPSSSPAAVDEPMVR

KMGKMTNLGNTLEARYEGSKSAAEYFDDKAAGYKVRIPVMDNTMTPTPTI

DKAFIDREOVVDWLSIRGFYIINSHDDMKIKEYNGNIEKRIWEOIARFNKS

ENLLEIMEPEPNTDEIDDMNSRLIKIIRKPTVILVIGGYWNSYNTLVNKKI

PPDPLIGTFHHYDPEFHKRGITMGTOEDDVTVRVDFPQGDMDSGTDLDDMS

AVMAADRTSPKWKYDFDLSDALEBGFACSVMDNCFGSDMDAIVNDPTPEPEI

LNALNPTGTPSPKSPPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPT

KIVSGKTMGEVSTGTGTVISLPDDWSKMLKISDIKSVDSAAEIKPMI

AKRSINGVDGHWYSTIPDSWKTIELPESSFRRLDYOPGQDMGSDTDLDDMS

IHFMYANNSKGFVVDNKLIGATSDPTSIKHGDLNPEDNAVNSTDLMLTKRYIKSL

ELGTSEOEKFKKADLNDKNDVSDTLILKRYILKAISEIPI"

sig_peptide

/note="cellulase signal peptide"

mat_peptide

303..2870

/product="cellulase"

BASE COUNT 1003 a 572 c 628 g 799 t

ORIGIN

Query Match 15.9%; Score 151.6; DB 1; Length 3002;

Best Local Similarity 54.1%; Pred. No. 8; Se-24;

Matches 332; Conservative 0; Mismatches 279; Indels 3; Gaps 1;

FEATURES

Location/Qualifiers

1..10415

/organism="Clostridium acetobutylicum"

1143 AAGGTGATGACCTTTTGAAGAACAAATTTGGGAAGAGGCACTTAATATAGGAAT 60

1143 AAGGTGATGACCTTTTGAAGAACAAATTTGGGAAGAGGCACTTAATATAGGAAT 1202

61 GCGCTTGAAGACCAATGAGGAGAGTGGTGTATTAAGATGAGTCTTCGCAC 120

1203 ACTCTCGAAGCTCCTATGAAGGCTCTGTCACAGTCCGATGGAATATTTTGAAT 1262

121 ATATATAAAGAGCGGTTCTCTCATGTTCAATTCGAATGAGTACGACGCT 180

1263 GATTTTAAAGTCGACAGATATAAAGGTAAGATCCCTGTAGATGGGCAACCATAC 1322

181 TACGCTTTCCTCCTTATAAATCATGATCGCTTCTCAAAAGAGTGGATGATGATA 240

1323 ATGAGAGACATACCCGATATACATTTGACAAAGCCTTTTGGACAGGGTTGACAGAGTGTT 1382

241 AAGGAGCCCTGAAAAGAGGAGTGGTGTGCTATTAATATTCATCATACGAGAGATTA 300

1383 GACTGTGACTTTTCAAGAGGTTTGTACAAATATATAATTCACATGATGACTGATC 1442

301 ATTAATGATTCAGAAACACAGCAAGAAAGATTCTTCTCTTGGAAACAATTCGCTGAT 360

1443 AAGGAAGACTATTAAGGAAACATAGAGGTTTGAAGAAGATATGAGGACGATTCGGGAA 1502

FEATURES

Location/Qualifiers

1..10415

/organism="Clostridium acetobutylicum"

1143 AAGGTGATGACCTTTTGAAGAACAAATTTGGGAAGAGGCACTTAATATAGGAAT 60

1143 AAGGTGATGACCTTTTGAAGAACAAATTTGGGAAGAGGCACTTAATATAGGAAT 1202

61 GCGCTTGAAGACCAATGAGGAGAGTGGTGTATTAAGATGAGTCTTCGCAC 120

1203 ACTCTCGAAGCTCCTATGAAGGCTCTGTCACAGTCCGATGGAATATTTTGAAT 1262

121 ATATATAAAGAGCGGTTCTCTCATGTTCAATTCGAATGAGTACGACGCT 180

1263 GATTTTAAAGTCGACAGATATAAAGGTAAGATCCCTGTAGATGGGCAACCATAC 1322

181 TACGCTTTCCTCCTTATAAATCATGATCGCTTCTCAAAAGAGTGGATGATGATA 240

1323 ATGAGAGACATACCCGATATACATTTGACAAAGCCTTTTGGACAGGGTTGACAGAGTGTT 1382

241 AAGGAGCCCTGAAAAGAGGAGTGGTGTGCTATTAATATTCATCATACGAGAGATTA 300

Db	398	TTCAAGAAAGTTGTCGATTATGCTCTTAACTGGTGGGTTATGCCATCTTAAACAATTCAC	457
Qy	287	ACTACG-----AGAGGTTAATGAATGATCCAGAGAACACAGAAAGATTCTTC	337
Db	458	ATGAACCTGGAAATTATGCTCTTCCAAAAGAAATTTAGAGAGTGCACAAAAGATCTTAGTTG	517
Qy	338	CTCTTTGGAAACAATTTCTGATTCGTTAAAGATCTCCGAACCTGATTTTGTGAA	397
Db	518	CCATCTGGAAACAATTCGCTGTAATTTGGTGAATTATGATGACACATTTATTTTGCAG	577
Qy	398	TTCTGAATGAACCTACGGAATCTTACTCCGGAATAATGGAATGAACCTGCTTGAGAA	457
Db	578	GAATGAATGAACCAAGAAAGTTGGGGATCCAGCTGAATGGACAGGTGATCAAGAAG	637
Qy	458	CTCTAAAGATTAAAGATCAATTGACAAAAGACACACTATATTTATAGGCACAGTGAA	517
Db	638	GTTGGAAATTCCTCAATGAATGAATGACCTTTTGTTAAACATATCTGCGCACTGGAG	697
Qy	518	G	518
Db	698	G	698

RESULT 14	AR130363	1826 bp	DNA	linear	PAT 16-MAY-2001
LOCUS	AR130363				
DEFINITION	Sequence	11 from patent	US 6190189.		
ACCESSION	AR130363				
VERSION	AR130363.1	GI:14118688			
KEYWORDS					
SOURCE	Unknown.				

REFERENCE	Unclassified.
AUTHORS	1 (bases 1 to 1826)
TITLE	Li, X.-L., Ljungdahl, L.G. and Chen, H.
JOURNAL	Cellulases and coding sequences
FEATURES	Patent: US 6190183-A 11 20-FEB-2001;
source	Location/Qualifiers
	1..1826

BASE COUNT	654 a	236 c	313 g	623 t
ORIGIN				

Query Match 7.78; Score 73; DB 6; Length 1826;

Best Local Similarity	30.4%	Fixed. NO. 3.36 00,
Matches 212; Conservative	0;	Mismatches 200; Indels 9; Gaps 1

107 ATGAGTCTTCGACATTATAAAGAGCGCTTCTCTCATGTTCCGATTCCCATTAAGAT 166

Db 277 AAGAGCTTACTATAAACTAAGTGATCTTGGTTTCACACACTTCCGTATTCCCTACTACTT 336

167 GGAGTACGACGGCTTACGGCTTCTCCTTATTAATCATGATCGCTTCTTCAAAAG 226

Db 337 GGAGTGGTCATTTTGGTGAITCCTCCIGACIAIAAAATAGTGAITCTTGGATGAAAGAG 350

[illegible]

337

457 ATGCAAGCTTTCGAAATTTATGCTCTTCTCCAAAGAAATTTAGAGAGTGGCCAAAGAGTCTTAGTTG 516

338 CTCTTTGGAACAATTGCTGATCGTTATTAAGACATATCCCGAAACCTCATTTTGTGAA 397

Db 517 CCATCTGGAACAAATTGCTGCTGAATTTGGTGATTATGATGAACATTTAAATTTTCGAAG 576

QY 398 TTCTGAATGAACCTCACGGAATCTTACTCCGGAATAATGGAATGAACCTTGAGGAG 457

Db 577 GAATGATGAACCAAGAAGGTTGGGATCCAGCTGATGGACAGGTGTGTATCAAGAAG 636

QY 458 CTCTAAAAGTTATAGATCAATTGACAAAAAGCACACTATAATTATAGGCACGCTGAAT 517

Db 637 GTGGATTTGCGCAATGAATGATGCGCCTTTTCGTTAAACTATTGTCGCACTGGAG 696
QY 518 G 518
Db 697 G 697

RESULT 15	LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS	SOURCE	ORGANISM
PE0277483	PE0277483	5502 bp mRNA for endoglucanase 5A (cel5A).	AJ277483	GI:8052315	cel5A gene; endoglucanase.	Pyromyces equi.	Pyromyces equi.

REFERENCE	AUTHORS	TITLE
1 (bases 1 to 5502)	Eberhardt, R. Y., Gilbert, H. J., and Hazlewood, G. P.	Primary sequence and enzymic properties of two modular

JOURNAL Microbiology 146 (Pt 8), 1999-2008 (2000)
 MEDLINE 20391845
 PUBMED 10931904
 2 (bases 1 to 5502)
 Eberhardt, R. Y.
 TITLE Direct Submission
 Submitted (20-APR-2000) Eberhardt R. Y., Molecular Enzymology, The
 Badriaham Institute, Babraham, Cambridge, CB2 4AT, UNITED KINGDOM
 FEATURES Location/Qualifiers

gene	103. .5247
	/gene="cel5A"
CDS	103. .5247
	"=15"

```

/codon_start=1
/evidence=experimental
/product="endoglucanase 5A"

```

```
/translation="MKFLNVLISITGLIYGSNAASSCSSEKLGKCEGKVVYIDN  
GKMGVENQKMGCIENEPPIVEPEPTTIVEPEPTTIVEPEPTTIVEPEPTT  
LEPTIDISSKELIKENFGNLTDAECTSMWNYEKDPGSEFTCMGNPKTIDM
```

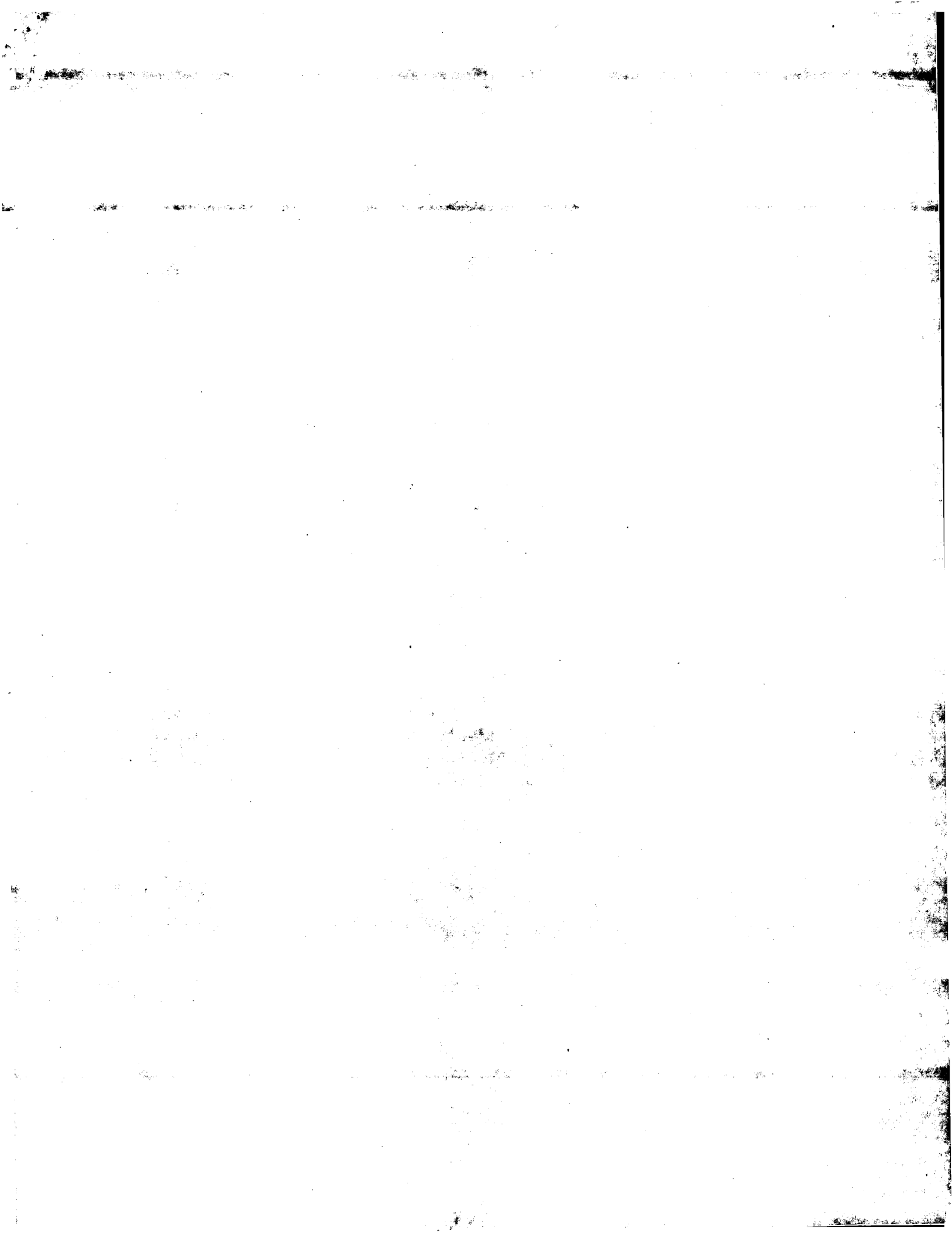
WDVVNMMNAVFMKTVRSSGGNNAKRHLMPYAAACNKNSTDFNDEPEDDDKVIASV
AYSPLYNALNNGEGAVDKFDATGKLELDYINGLIKKRFVSKGIPVIMGEYCAMNRDN
EVRATWAEIYMKETALGVBPVWMDNGIFEEGEGRGLIDBKNLKVYPSIYALQX

LORNDON ILEBES IORNDON
 GHIEAPDYK INEKMAKRVHEI VDY PY KNGAFVILNI HHESWNHAEETVEEAKVELL
 KWMAQ IAEERKDYDEHL IFEGONEPRKNDTPEWNGGDQEGMDVYANAMAVFMKTYR
 SGGNAKRHLMI PPYAAACNONSFDHEDFEPDDDKXI ASYAHY SPINFALNNGEGAV

[illegible]

NSFDNEDEPEDDCKVLASVHAISFIALNNOGSHVNDLQAOONNLSINCEEN
VSKGIPYIMGEIGAMNDRNEBRATAYKKEITLALGIPQWMMDDCIFEEGGERG
IDRKNLKVIPSIVALAQKRGSLVAVLHALEPKPEPPTTVVEEBETLAVDEPIS
YHDPJOURDTSCKETAEEMENENANTIDACTSMWNYEKDOPISGTCWGNPKTTE

HESMNHAEIJEAEAKVELAKWAUJAEFEEDI DEHIE EGVNFEINNDI V UNNO
 QECMDVYNAMAVFMKIVRSSGGNNAKRHLMTPIYAACNKSFNDFPEDDDKV
 SVHAYSPYNFALNNGEAVDKFDATGCHNELLDJNLGLIKKRFVSKGIVIMBEYGAM



GenCore version 5.1.4-p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 26, 2003, 07:29:37 ; Search time 191 Seconds
(without alignments)
11248.201 Million cell updates/sec

Title: US-09-880-729A-1

Perfect score: 954

Sequence: 1 atgggtgttgatccttga.....gagagatagcatgtaataa 954

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 112599159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N.Geneseq_101002.*

1: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.*
2: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.*
3: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.*
4: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.*
5: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.*
6: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.*
7: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.*
8: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.*
9: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.*
10: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.*
11: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.*
12: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.*
13: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.*
14: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.*
15: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.*
16: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.*
17: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.*
18: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.*
19: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.*
20: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.*
21: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.*
22: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.*
23: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.*
24: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	954	100.0	954	20	AAZ21563
2	954	100.0	954	20	AAK85879
3	954	100.0	954	21	AAZ46759
4	954	100.0	954	22	AAAD15243
5	235.8	24.7	1083	19	AAV3619
6	232.6	24.4	1083	18	AAV3690
7	73	7.7	1826	19	AAV29477
8	73	7.7	1826	19	AAV37413
9	68.6	7.2	1532	21	AAA64503

10	63.2	6.6	4098	15	AAQ55036
11	62.8	6.6	1170	22	AAE83293
12	62.8	6.6	1182	22	AAE83296
13	62.8	6.6	2110	22	AAE83292
14	49.4	5.2	1614	20	AAK05950
15	48.2	5.1	2777	17	AAK41849
16	42.6	4.5	32184	22	ABA20589
17	42.6	4.5	32184	22	ABA20590
18	42.6	4.5	38272	22	AAK84952
19	42.6	4.5	38348	22	AAK84953
20	42.4	4.4	6467	24	ABN80169
21	42	4.4	2172	20	AAK61594
22	42	4.4	2262	20	AAK61593
23	41.6	4.4	3740	15	AAQ54682
24	40.8	4.3	6290	24	ABL33046
25	40.6	4.3	7442	22	AAK46686
26	40.2	4.2	585	24	ABO60399
27	40.2	4.2	7317	22	AAK45342
28	40.2	4.2	7317	24	ABK28173
29	40.2	4.2	16228	24	ABL70459
30	40.2	4.2	16228	24	AAK61424
31	39.8	4.2	6392	24	ABL32685
32	39.8	4.2	6392	24	ABL34507
33	39.2	4.1	2166	11	AAQ06781
34	38.8	4.1	6853	24	ABL54331
35	38.8	4.1	6853	24	ABL32310
36	38.2	4.0	5501	24	ABL32054
37	38	4.0	2623	23	ABL12172
38	38	4.0	4288	23	ABN05576
39	38	4.0	8067	24	ABN80224
40	38	4.0	580073	18	AAK58840
41	37.8	4.0	1092	22	AAAD16819
42	37.8	4.0	1284	23	AAK70787
43	37.8	4.0	6740	24	ABL33147
44	37.6	3.9	611	23	ABV51437
45	37.6	3.9	6863	22	AAK46618

ALIGNMENTS

RESULT 1	AAZ21563
ID	AAZ21563 strand; cDNA: 954 BP.
XX	AAZ21563;
AC	AAZ21563;
XX	XX
DT	01-DEC-1999 (first entry)
XX	XX
DE	Carboxymethyl cellulase (CMCase) gene.
XX	XX
KW	Carboxymethyl cellulase; CMCase; thermostable enzyme; hydrolysis;
KW	beta 1 4 glycosidic bond; cellulose; plant biomass conversion; fuel;
KW	chemical; detergent; textile industry; animal feed; waste treatment;
KW	fruit juice extraction; clarification; brewing industry; ss.
OS	Thermotoga maritima.
XX	XX
FH	Key
FT	CDS
FT	Location/Qualifiers
FT	1..954
FT	/*tag= a
FT	/product= CMCase
FT	/note="Carboxymethyl cellulase"
XX	XX
PN	US5962258-A.
PD	05-OCT-1999.
XX	XX
PF	23-AUG-1995; 95US-0518615.
XX	XX
PR	23-AUG-1995; 95US-0518615.
XX	XX
PA	(DIVE-) DIVERSA CORP.

Sequence of celd c
P. rhizinflata egl
P. rhizinflata egl
P. rhizinflata egl
Bacillus agardhae
Bacillus cellulase
Human nervous syst
Human nervous syst
Human immune/haema
Human immune/haema
Human chemically m
B. burgdorferi ant
Potato sucrose pho
Human immune syste
Tumour suppressor
Human colon cancer
Chemically pretrea
DNA transcription
Chemically treated
Human gene regulat
Human immune syste
Human metastasis a
Sequence encoding
Chemically treated
Human immune syste
Human immune syste
Drosophila melanog
Drosophila melanog
Human chemically m
Mycoplasma genital
Paenibacillus pabu
DNA encoding novel
Human immune syste
Human prostate exp
Tumour suppressor

XX Mathur EJ, Lam DE;
 XX WPI: 1999-571269/48.
 DR P-PSDB; AAY39338.
 PT Recombinant polynucleotide sequence encoding a carboxymethyl cellulase
 XX useful for the conversion of plant biomass -
 PS Claim 11; Fig 1; 14pp; English.

This sequence is a carboxymethyl cellulase gene (CMCase) CMCase is a
 CC thermostable enzyme (optimum temperature of 60 + degrees C) that
 CC catalyses the hydrolysis of the beta 1,4 glycosidic bonds in cellulose.
 CC The enzyme degrades cellulose by the hydrolysis of this bond. The CMCase
 CC enzyme has a molecular weight of approximately 35KD. CMCase is used to
 CC catalyse enzymatic degradation of cellulose and can be used for the
 CC conversion of plant biomass into fuels and chemicals. CMCase can also be
 CC used in detergents, the textile industry, animal feed, waste treatment
 CC and for the clarification and extraction of juices in the fruit juice and
 CC brewing industries. The CMCase gene and 15 bp fragments of it can be used
 CC as probes to isolate nucleic acid molecules encoding CMCase enzymes from
 CC genomic libraries or to screen for related sequences. Cells expressing
 CC the CMCase polynucleotide can be used as immunogens to produce antibodies
 CC which can then be used to isolate the enzyme from cells which express
 CC CMCase. The antibodies can also be used to screen for similar enzymes
 CC from other organisms and samples.

Sequence 954 BP; 335 A; 145 C; 221 G; 253 T; 0 other;

Query Match 100.0%; Score 954; DB 20; Length 954;

Best Local Similarity 100.0%; Pred. No. 1.6e-244;
 Matches 954; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGGGTGTTGATCTTTTGAAGAACAAATATTTGGGAAGCGCTTATATAGCAAT 60
 DB 1 ATGGGTGTTGATCTTTTGAAGAACAAATATTTGGGAAGCGCTTATATAGCAAT 60
 QY 61 GCGCTTGAAGCAACCAATAGGAGAGCTGGGGAGTGTGTAAATGATGCTTCTCGAC 120
 DB 61 GCGCTTGAAGCAACCAATAGGAGAGCTGGGGAGTGTGTAAATGATGCTTCTCGAC 120
 QY 121 ATTATTAAGAAAGCCGCTTCTCATGTTTCAATTCATAAGATGAGTACGACGCT 180
 DB 121 ATTATTAAGAAAGCCGCTTCTCATGTTTCAATTCATAAGATGAGTACGACGCT 180
 QY 181 TACGCGTTCTCTTATTAATATCATGATCGCTTCTTCAAAAGAGTGAAGTAT 240
 DB 181 TACGCGTTCTCTTATTAATATCATGATCGCTTCTTCAAAAGAGTGAAGTAT 240
 QY 241 AACGGAGCCCTGAAAAGAGAGCTGGCTGTCTATAATATTCATCACTCGAGAGTTA 300
 DB 241 AACGGAGCCCTGAAAAGAGAGCTGGCTGTCTATAATATTCATCACTCGAGAGTTA 300
 QY 301 ATGAATGATCCAGAGAACCAAGAAAGATTTCTTGGAAACAAATTTGCTGAT 360
 DB 301 ATGAATGATCCAGAGAACCAAGAAAGATTTCTTGGAAACAAATTTGCTGAT 360
 QY 361 CGTTATTAAGACATATCCCGAACTATTTTGAATTTCTGAATGAACTCAGCAAT 420
 DB 361 CGTTATTAAGACATATCCCGAACTATTTTGAATTTCTGAATGAACTCAGCAAT 420
 QY 421 CTACTCCGGAAGAAATGATGAACTGTTGAGAACTCTAAAGTTATTAAGTCAATT 480
 DB 421 CTACTCCGGAAGAAATGATGAACTGTTGAGAACTCTAAAGTTATTAAGTCAATT 480
 QY 481 GACAAAAAGCACACTATATATATAGGACAGCTAATGGGGGTATATCTGCCCTTGA 540
 DB 481 GACAAAAAGCACACTATATATATAGGACAGCTAATGGGGGTATATCTGCCCTTGA 540
 QY 541 AAACGTGCTGCCCAAAATGGGAAATATCTATAGTACAAATCTACTACATCTCT 600
 DB 541 AAACGTGCTGCCCAAAATGGGAAATATCTATAGTACAAATCTACTACATCTCT 600

QY 601 TTGCAATTTACCCATCAAGAGCTGAGTGGTGAAGGATCTGAAATGTTGGAGA 660
 DB 601 TTGCAATTTACCCATCAAGAGCTGAGTGGTGAAGGATCTGAAATGTTGGAGA 660
 QY 661 AAGTGGGATCTCCAGATGATCAAGAACATTTGATAGAACATTTATATAGAGA 720
 DB 661 AAGTGGGATCTCCAGATGATCAAGAACATTTGATAGAACATTTATATAGAGA 720
 QY 721 TGTGCAAAAAGAACCAAAAGCAATTTACATAGGTAGTGTGCTACAGAAAGCT 780
 DB 721 TGTGCAAAAAGAACCAAAAGCAATTTACATAGGTAGTGTGCTACAGAAAGCT 780
 QY 781 GACCTGATCAATCAATTAATGACCTCTTGTCTTCCGCAAAATGAGAAAGAGA 840
 DB 781 GACCTGATCAATCAATTAATGACCTCTTGTCTTCCGCAAAATGAGAAAGAGA 840
 QY 841 TGAAGCTGGGCACTCTGGGAATTTGTCGGGTTTGGTGTATATGATCTGAGAAA 900
 DB 841 TGAAGCTGGGCACTCTGGGAATTTGTCGGGTTTGGTGTATATGATCTGAGAAA 900
 QY 901 ACCTGATTAAGATCTTTTGAAGCTTTAATAGAGAGATAGCATTAATAA 954
 DB 901 ACCTGATTAAGATCTTTTGAAGCTTTAATAGAGAGATAGCATTAATAA 954

RESULT 2

AAK85879 standard; cDNA; 954 BP.

AAK85879;

10-SEP-1999 (first entry)

CDNA encoding a carboxymethyl cellulase enzyme.

Carboxymethyl cellulase enzyme: hydrolysis; cellulose; plant biomass;

Fuel; Chemical; detergent; textile industry; animal feed;

Waste treatment; fruit juice; brewing industry; clarification;

juice extraction; ss.

Thermotoga maritima.

US5925749-A.

20-JUL-1999.

23-AUG-1995; 95US-0518615.

23-AUG-1995; 95US-0518615.

24-APR-1996; 98US-0066075.

(DIVE-) DIVERSA CORP.

Lam DE, Mathur EJ;

WPI: 1999-429058/36.

P-PSDB; AAY23764.

New carboxymethyl cellulase enzyme is useful in cellulose digestion

Claim 1; Fig 1A-C; 13pp; English.

The present sequence encodes a carboxymethyl cellulase enzyme from
 CC Thermotoga maritima. Oligonucleotide probes derived from the
 CC polynucleotide sequence can be used to screen a library of genomic
 CC DNA for recovery of the polynucleotide, as PCR primers to synthesize
 CC full length polynucleotides or as diagnostic probes useful for
 CC screening for similar enzymes from other organisms and samples.
 CC The carboxymethyl cellulase enzymes can be used to catalyse hydrolysis
 CC of cellulose for the conversion of plant biomass into fuels and
 CC chemicals, for use in detergents, the textile industry, in animal feed,
 CC waste treatment and in the fruit juice/brewing industry for the

CC clarification and extraction of juices.

XX Sequence 954 BP; 335 A; 145 C; 221 G; 253 T; 0 other;

Query Match 100.0%; Score 954; DB 20; Length 954;
Best Local Similarity 100.0%; Pred. No. 1.6e-244;
Matches 954; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGGGTTGATCCTTTGAAAGCAAAATTTGGGAAGGCACTTAATATAGCAAAAT 60
DB 1 ATGGGTTGATCCTTTGAAAGCAAAATTTGGGAAGGCACTTAATATAGCAAAAT 60
QY 61 GCGCTTGAAGCACCACCAATGAGGAGACTGGGAGTGTGATTAAGAATGAGTCTTGAC 120
DB 61 GCGCTTGAAGCACCACCAATGAGGAGACTGGGAGTGTGATTAAGAATGAGTCTTGAC 120
QY 121 ATTATTAAGAAGCCGGTTTCTCATGTTGCAATTCGAATTAAGATGAGTACGACGCT 180
DB 121 ATTATTAAGAAGCCGGTTTCTCATGTTGCAATTCGAATTAAGATGAGTACGACGCT 180
QY 181 TACGCTTTCCTCTTATAAATCATGATCGCTTCTCAAAAGAGTGAATGATGATA 240
DB 181 TACGCTTTCCTCTTATAAATCATGATCGCTTCTCAAAAGAGTGAATGATGATA 240
QY 241 AACGAGCCCTGAAAAGAGACTGGCTGTTGCTATAAATATTCATCACTACGAGAGTTA 300
DB 241 AACGAGCCCTGAAAAGAGACTGGCTGTTGCTATAAATATTCATCACTACGAGAGTTA 300
QY 301 ATGAAATGATCCAGAACACCAAGAAAGATTTCTGCTTTGGAAACAAATGCTGAT 360
DB 301 ATGAAATGATCCAGAACACCAAGAAAGATTTCTGCTTTGGAAACAAATGCTGAT 360
QY 361 CTTTATAAGACTATCCCGAACTATTTTGAATTCGATGATACCTCAGCAAT 420
DB 361 CTTTATAAGACTATCCCGAACTATTTTGAATTCGATGATACCTCAGCAAT 420
QY 421 CTTTACTCGGAAAAATGGAATGAACTGTTGAGAACTCTAAGATTAAGATCAAT 480
DB 421 CTTTACTCGGAAAAATGGAATGAACTGTTGAGAACTCTAAGATTAAGATCAAT 480
QY 481 GACAAAACACACTATTAATATAGGCACAGCTGATGGGGGTATATCTGCCCTTGA 540
DB 481 GACAAAACACACTATTAATATAGGCACAGCTGATGGGGGTATATCTGCCCTTGA 540
QY 541 AAACGTCTGTCCCAAAATGAGAAAAATTTCTATAGTTACAAATCACTACTCAATCC 600
DB 541 AAACGTCTGTCCCAAAATGAGAAAAATTTCTATAGTTACAAATCACTACTCAATCC 600
QY 601 TTGCAATTTACCATCAAGAGAGCTGAGTGGGTGGAAGATCTGAGAAATGTTGGGA 660
DB 601 TTGCAATTTACCATCAAGAGAGCTGAGTGGGTGGAAGATCTGAGAAATGTTGGGA 660
QY 661 AAGTGGGATCTCCATGATGTCAGAAACATTTGATAGAAATTCATTTATAGAGAA 720
DB 661 AAGTGGGATCTCCATGATGTCAGAAACATTTGATAGAAATTCATTTATAGAGAA 720
QY 721 TGGTCAAAAAAGAACCAATTTTACATAGTGTGAGTTGGTCCCTACAGAAAGCT 780
DB 721 TGGTCAAAAAAGAACCAATTTTACATAGTGTGAGTTGGTCCCTACAGAAAGCT 780
QY 781 GACCTTGAATCAAGAAATTAATGAGCTCTTGTCTGCGGAATGAGAAAGAGAGA 840
DB 781 GACCTTGAATCAAGAAATTAATGAGCTCTTGTCTGCGGAATGAGAAAGAGAGA 840
QY 841 TGGAGCTGGGCAATCTGGAATTTTGTCCGGTTTGGTGTATATCTCGAAGAAA 900
DB 841 TGGAGCTGGGCAATCTGGAATTTTGTCCGGTTTGGTGTATATCTCGAAGAAA 900
QY 901 ACCTGGAATTAAGATCTTTAGAGCTTTAATAGAGAGATGACATTAATA 954
DB 901 ACCTGGAATTAAGATCTTTAGAGCTTTAATAGAGAGATGACATTAATA 954

RESULT 3
AAZ46759
ID AAZ46759 standard; cDNA; 954 BP.

XX AC AAZ46759;

DT 31-MAR-2000 (first entry)

DE T. maritima CMCase enzyme encoding cDNA.

KW Endoglucanase; carboxymethyl cellulase; cellulose; CMCase; plant biomass;
KW beta-1,4-glycosidic bond; detergent; textile; guar gum; animal feed;
KW waste product; fuel; brewing; thermostable; ss.

OS Thermotoga maritima.

FT Key Location/Qualifiers
FT CDS 1..954
FT /tag- a
FT /product- "CMCase"

PN US6008032-A.

PD 28-DEC-1999.

PF 16-OCT-1997; 97US-0951889.

PR 23-AUG-1995; 95US-0518615.

PA (DIVE-) DIVERSA CORP.

PI Mathur EJ, Lam DE;

DR WPI: 2000-105335/09.

DR P-PSDB; AAY56814.

XX

PT Endoglucanase with carboxymethyl cellulase activity from Thermotoga
PT maritima, useful for degrading biomass to fuels and chemicals

XX

PS

XX This cDNA encodes a T. maritima endoglucanase with carboxymethyl
XX cellulase activity. The enzyme CMCase is thermostable and used to degrade
XX carboxymethyl cellulose (CMC) or more generally any oligosaccharide that
XX contains beta-1,4-glycosidic bonds, particularly in detergents, textiles,
XX guar gum, animal feed, plant biomass or waste products. The CMCase enzyme
XX is especially used to convert plant biomass to chemicals and fuels, and
XX also for clarification and extraction of juice (e.g. In the brewing
XX industry). CMCase can also be used to raise specific antibodies.

Sequence 954 BP; 335 A; 145 C; 221 G; 253 T; 0 other;

Query Match 100.0%; Score 954; DB 21; Length 954;
Best Local Similarity 100.0%; Pred. No. 1.6e-244;
Matches 954; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGGGTTGATCCTTTGAAAGCAAAATTTGGGAAGGCACTTAATATAGCAAAAT 60

DB 1 ATGGGTTGATCCTTTGAAAGCAAAATTTGGGAAGGCACTTAATATAGCAAAAT 60

QY 61 GCGCTTGAAGCACCACCAATGAGGAGACTGGGAGTGTGATTAAGAATGAGTCTTGAC 120

DB 61 GCGCTTGAAGCACCACCAATGAGGAGACTGGGAGTGTGATTAAGAATGAGTCTTGAC 120

QY 121 ATTATTAAGAAGCCGGTTTCTCATGTTGCAATTCGAATTAAGATGAGTACGACGCT 180

DB 121 ATTATTAAGAAGCCGGTTTCTCATGTTGCAATTCGAATTAAGATGAGTACGACGCT 180

QY 181 TACGCTTTCCTCTTATAAATCATGATCGCTTCTTCAAAAGAGTGAATGATGATA 240

DB 181 TACGCTTTCCTCTTATAAATCATGATCGCTTCTTCAAAAGAGTGAATGATGATA 240

QY 241 AACGAGCCCTGAAAAGAGACTGGCTGTTGCTATAAATATTCATCACTACGAGAGTTA 300

Db	241	AACGGAGCCCTGAAAAAGAGGAGCTGGCTGTCTCTAATAATATTCACTACAGAGAGTTA	300
Qy	301	ATGAATGATCCAGAGAAGAACACAGAGAAAGATTTCCTGCTCTTTGGAAACAAATTGCTGAT	360
Db	301	ATGAATGATCCAGAGAAGAACACAGAGAAAGATTTCCTGCTCTTTGGAAACAAATTGCTGAT	360
Qy	361	CGTTATTAAGACTATCCCGGAAGCTATTTTTGGAATTCGAATGAACTCAGAGGAAT	420
Db	361	CGTTATTAAGACTATCCCGGAAGCTATTTTTGGAATTCGAATGAACTCAGAGGAAT	420
Qy	421	CTTACTCCGGAAAAATGGAATGAACCTGCTTGAAGAGCTCTPAAAGTTATTAAGTCAATT	480
Db	421	CTTACTCCGGAAAAATGGAATGAACCTGCTTGAAGAGCTCTPAAAGTTATTAAGTCAATT	480
Qy	481	GACAAAAGACACACTATTAATTATAGGCGACAGCTGAAATGGGGGGGATATCTGCCCTTGA	540
Db	481	GACAAAAGACACACTATTAATTATAGGCGACAGCTGAAATGGGGGGGATATCTGCCCTTGA	540
Qy	541	AAACTGTCTGTCCCAAAATGGGAAAAAAATTCATATAGTTACAAATTCACCTACTACAACTCT	600
Db	541	AAACTGTCTGTCCCAAAATGGGAAAAAAATTCATATAGTTACAAATTCACCTACTACAACTCT	600
Qy	601	TTTCAATTTTACCCATCAAGAGAGCTAGTGGGTGGAGAAATCTGAGAAATGGTTGGAGAA	660
Db	601	TTTCAATTTTACCCATCAAGAGAGCTAGTGGGTGGAGAAATCTGAGAAATGGTTGGAGAA	660
Qy	661	AAGTGGGATCTCCAGATGATCAGAAACATTTGATAGAGAAATTCATATTTTATAGAGAA	720
Db	661	AAGTGGGATCTCCAGATGATCAGAAACATTTGATAGAGAAATTCATATTTTATAGAGAA	720
Qy	721	TGCTCAAAAAAGACAAAGACCAATTTTACATAGTGAATTGGTTGGTCCCTACAGAAAGCT	780
Db	721	TGCTCAAAAAAGACAAAGACCAATTTTACATAGTGAATTGGTTGGTCCCTACAGAAAGCT	780
Qy	781	GACCTTGAATCAAGATTAATAATGGAACCTCTTGTCTGCTCGGAAATGAGAGAAAGAGA	840
Db	781	GACCTTGAATCAAGATTAATAATGGAACCTCTTGTCTGCTCGGAAATGAGAGAAAGAGA	840
Qy	841	TGGAGCTGGGCACTACTGGGAATTTTGTCCGCTTTTGGTGTATTATGATACCTGAGAAAA	900
Db	841	TGGAGCTGGGCACTACTGGGAATTTTGTCCGCTTTTGGTGTATTATGATACCTGAGAAAA	900
Qy	901	ACCTGGAATTAAGATCTTTTATAGGCTTAAATAGGAGAGATAGCATTAATTA	954
Db	901	ACCTGGAATTAAGATCTTTTATAGGCTTAAATAGGAGAGATAGCATTAATTA	954
RESULT 4			
AADI5243			
ID	AADI5243	standard; DNA; 954 BP.	
XX	AADI5243;		
XX	AADI5243;		
XX	15-NOV-2001 (first entry)		
DE	Thermotoga maritima CMCase enzyme encoding DNA.		
XX			
KM	Carboxymethyl cellulose; CMCase; plant biomass; animal feed; detergent;		
KM	textile industry; waste treatment; brewing industry; immunogen;		
KW	fruit juice industry; ds.		
XX			
OS	Thermotoga maritima.		
XX			
PH	Key	Location/Qualifiers	
FT	CDS	1..954	
FT	/*tag= "a		
FT	/product= "T. maritima CMCase enzyme "		
XX			
PN	US6245547-B1.		
XX			
PD	12-JUN-2001.		
XX			

PF	27-DEC-1999.	99US-0472857.
XX	23-AUG-1995.	95US-0518615.
PR	16-OCT-1997.	97US-0951889.
XX	(DIVE-) DIVERSA CORP.	
PA	Mathur EJ, Iam DE;	
PI		
XX	WPI: 2001-528380/58.	
DR	P-PSDB; AAE08543.	
XX		
PT	Polynucleotide isolated from thermophile eubacteria useful for	
PT	degradation of cellulose encodes an enzyme having a fully defined amino	
XX	acid sequence -	
XX		
PS	Claim 3; Fig 1; 14pp; English.	
XX		
CC	The present sequence is a DNA coding for a thermostable enzyme, an	
CC	endoglucanase having carboxymethyl cellulose activity referred to as	
CC	(CMCase). The CMCase enzyme is useful for the degradation of cellulose,	
CC	for conversion of plant biomass into fuels and chemicals, in the	
CC	textile industry, in detergents, in animal feed, in waste treatment,	
CC	in fruit juice/brewing industry for clarification and extraction of	
CC	juices, to generate probes for identifying sequences encoding similar	
CC	enzymes from other organisms, as immunogen to produce monoclonal or	
XX	polyclonal antibodies.	
XX		
SQ	Sequence 954 BP; 335 A; 145 C; 221 G; 253 T; 0 other:	
	Query Match 100.0%; Score 954; DB 22; Length 954;	
	Best Local Similarity 100.0%; Pred. No.1,6e-244;	
	Matches 954; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
OY	1 ATGGGTTGTATGCCCTTGTGAAGAAACAATAATTGGCAAGAAGCATTAATATAGCAAT	60
DB	1 ATGGGTTGTATGCCCTTGTGAAGAAACAATAATTGGCAAGAAGCATTAATATAGCAAT	60
OY	61 GCGGTTGAAGACCAAAATGAGGAGACTGGGAGTGGTGATAAAAAGATGATTCCTTGAC	120
DB	61 GCGGTTGAAGACCAAAATGAGGAGACTGGGAGTGGTGATAAAAAGATGATTCCTTGAC	120
OY	121 ATTATAAAAGAGCGGTTTTCTCTCATGTGCATTTCCATTAAGATGAGTAGCAGCT	180
DB	121 ATTATAAAAGAGCGGTTTTCTCTCATGTGCATTTCCATTAAGATGAGTAGCAGCT	180
OY	181 TAGCGGTTCTCTCTTTAAATCATGATATCGCTTCCAAGAGAGTGATGAGAGTGATA	240
DB	181 TAGCGGTTCTCTCTTTAAATCATGATATCGCTTCCAAGAGAGTGATGAGAGTGATA	240
OY	241 AACGAGGCGCTGAANAAGAGAGCTGGCTGTTCATTAATATATTCATCTACGAGAGTTA	300
DB	241 AACGAGGCGCTGAANAAGAGAGCTGGCTGTTCATTAATATATTCATCTACGAGAGTTA	300
OY	301 ATGAATGATCCAGAAGAACACAAAGGAAGATTTCTGGCTTTGGAAACAATTGCGAT	360
DB	301 ATGAATGATCCAGAAGAACACAAAGGAAGATTTCTGGCTTTGGAAACAATTGCGAT	360
OY	361 CGTTATTAAGACTATCCCGAAGACTCTATTTTTTGAATTCGTAATGAAGACCTACGGAAT	420
DB	361 CGTTATTAAGACTATCCCGAAGACTCTATTTTTTGAATTCGTAATGAAGACCTACGGAAT	420
OY	421 CTTACTCCGGAAAAATGGAATGAAGTCTGTGAGGAAGCTCTTAAAGTTATTAAGTCAATT	480
DB	421 CTTACTCCGGAAAAATGGAATGAAGTCTGTGAGGAAGCTCTTAAAGTTATTAAGTCAATT	480
OY	481 GACAAAAGACACTATTAATTAATAGCACGCGTAATGGGGGGGTATATCGCCCTTGA	540
DB	481 GACAAAAGACACTATTAATTAATAGCACGCGTAATGGGGGGGTATATCGCCCTTGA	540
OY	541 AAACGTCTGTGCCCAAATGGGAAAAAATTTCTAAGTTACATTCACCTACTACATCTCT	600
DB	541 AAACGTCTGTGCCCAAATGGGAAAAAATTTCTAAGTTACATTCACCTACTACATCTCT	600

QY 601 TTGCAATTTTCCCATCAAGAGCTGAGTGGTGGAGATCTGAGAAATGTTGGAGAGA 660
 DB 601 TTGCAATTTTCCCATCAAGAGCTGAGTGGTGGAGATCTGAGAAATGTTGGAGAGA 660
 QY 661 AAGTGGGATCTCCAGATGATCAGAAACATTTGATAGAGAAATTTTATAGAGAA 720
 DB 661 AAGTGGGATCTCCAGATGATCAGAAACATTTGATAGAGAAATTTTATAGAGAA 720
 QY 721 TGTGCAAAAAGAACAAACCAATTTACATAGTGTGAGTGGTGGCTCAGAAAAGCT 780
 DB 721 TGTGCAAAAAGAACAAACCAATTTACATAGTGTGAGTGGTGGCTCAGAAAAGCT 780
 QY 781 GACCTTGAATCAAGATTAATGAGACCTCTTGTCTGCGCAAAATGAGAAAAGAGA 840
 DB 781 GACCTTGAATCAAGATTAATGAGACCTCTTGTCTGCGCAAAATGAGAAAAGAGA 840
 QY 841 TGGAGCTGGGATCTGAGATTTTGTCCGTTTGGTGTATGATCTGAGAGAA 900
 DB 841 TGGAGCTGGGATCTGAGATTTTGTCCGTTTGGTGTATGATCTGAGAGAA 900
 QY 901 ACCTGGAATTAAGATCTTTTGAAGCTTTAATGAGAGATGACATGATAA 954
 DB 901 ACCTGGAATTAAGATCTTTTGAAGCTTTAATGAGAGATGACATGATAA 954

RESULT 5
 AAV36919
 ID AAV36919 standard; DNA: 1083 BP.
 AC AAV36919;
 DT 21-DEC-1998 (first entry)
 DE Thermotoga OCl/4V endoglucanase gene coding region.
 XX Glycosidase; OCl/4V; thermostable enzyme; oligosaccharide; glucose;
 KM sugar; baking; textile; detergent; endoglucanase; ss.
 OS Thermotoga sp. strain OCl/4V-33GPI.
 PN WO9824799-A1.
 PD 11-JUN-1998.
 PE 08-DEC-1997; 97MO-US22623.
 PR 10-OCT-1997; 97US-0949026.
 PR 06-DEC-1996; 96US-0056916.
 PA (DIVE-) DIVERSA CORP.
 PI Bylina EJ, Lam DE, Mathur EJ, Swanson RV;
 DR WPI; 1998-362407/31.
 DR P-PSDB; AAW49870.
 PT Glycosidase enzymes from organisms of the genera *Staphylothermus*,
 PT *Pyrococcus* and *Thermococcus* - for deriving sugar from
 PT oligosaccharides, useful in the e.g. food processing, textile or
 PT baking industries
 XX Claim 1; Fig 13a-b; 92pp; English.
 CC This isolated polynucleotide comprises a coding region for
 CC endoglucanase OCl/4V (see AAW49870) from a *Thermotoga* sp. (clone
 CC OCl/4V-33GPI) that grows optimally at 75 degC. The sequence
 CC shows 43% nucleic acid identity to an endo-1,4-beta-endoglucanase
 CC of *Caldocellum* thermocellum. The invention provides 18
 CC polynucleotides (see AAV36907-24) coding for thermostable glycosidases
 CC (see AAW49888-75) having glucosidase, alpha-galactosidase,
 CC beta-galactosidase, beta-mannosidase, beta-mannanase, endoglucanase
 CC or pullulanase activity. Vectors and host cells are also claimed.

CC A method is provided for producing the enzymes by recombinant
 CC techniques. A claimed method for generating glucose from soluble
 CC cell oligosaccharides comprises contacting a sample (selected from
 CC dairy products, fruit juice, detergent, textile, guar gum, animal
 CC feed, plant biomass or waste product) containing oligosaccharides
 CC (selected from maltose, cellobiose, lactose, sucrose, raffinose,
 CC stachyose, verbasinose, cellulose, starch, amylose, glycogen,
 CC disaccharides, polysaccharides and pullulan) with one of the
 CC claimed glycosidases such that glucose is produced.
 XX
 SQ Sequence 1083 BP; 367 A; 189 C; 242 G; 285 T; 0 other;
 Query Match 24.7%; Score 235.8; DB 19; Length 1083;
 Best Local Similarity 54.2%; Pred. No. 6,5e-53;
 Matches 502; Conservative 0; Mismatches 422; Indels 3; Gaps 1;

QY 14 CTTTGAAGAGACAAATATTTGGAGAGGACATTAATATAGAAATGCGCTTGAAGCAC 73
 DB 149 CATTGGAATACAAACAAATATGATAGTAAGAGATTAATATGGAATGCTTGAAGACCTC 208
 QY 74 CAATGAGGGAGACTGGGAGTGTGATAAAGATGAGTCTTGCACATTAATAAAGAG 133
 DB 209 CTTTGAAGAGACCTTGGGAGTAAGATTTGAGATGATATTTTGAATATTAAGAGAAA 268
 QY 134 CCGGTTTCTCATGTTGCAATTCATTAAGATGAGTACGACGCTTACGCGTTTCTC 193
 DB 269 GGGGATTTGATTTCTGTAGAGATTTCCCATTAAGATGTCAGACATATATCCGAAAGGCAC 328
 QY 194 CTTATTAATCATGATGATGCTTCTTCAAAAGATGATGATGATTAAGGAGCCTGA 253
 DB 329 CATATGATATTGACAGGAATTTCTCCGAAAGATTAACATATGTTGTGATAGGCTCTTG 388
 QY 254 AAGAGAGACTGGCTGTGCTATTAATATATCATATCATGAGGAGATTAATGATGATCAG 313
 DB 389 AGATTAATTTTACAGTATATCATATCATATGACACCTTTTGAAGAACTATACAGAACCGG 448
 QY 314 AAGAACCAAGAGAAAGATTTCTGCTCTTTGAAACAAATTTGCTGATGTTATTAAGACT 373
 DB 449 ATAAATGAGGAGATTTTGTGTGTAATTTGAGACAGATTTGCAAAATTTCTTAAAGATT 508
 QY 374 ATCCGGAACCTATTTTGAATTTCTGAATGACCTCAGGAAATCTTACTCGGAAA 433
 DB 509 ACCCGGAAATCTGTTTGAATTTCTGAAATCTACAAAGCAGCTCTCAGAACTTGACACTTAAA 568
 QY 434 AATGGAATGACGCTTGAGAGACCTCTAAAGCTTATTAAGATCATTTGACAAAGACACA 493
 DB 569 AATGGAACGACATTTATTCAAAGATGCTCAAGATTTACAGGAGAGCAATTCACACCGGA 628
 QY 494 CTATTAATTTAGGACACAGTGAATGAGGGGGTATATCTGCCCTTGAAAAACCTGTCTGCC 553
 DB 629 TTGTCATTTATGATGCTCCAAACCTGGGACACATATACGCGAGTGAAGAGCTTAATTAAG 688
 QY 554 CAAATGAGGAAAAATTTCTATAGTTCATCTACATCTACATCTTTTGAATTTTACC 613
 DB 689 TCAA--CGACAAACGACATATGTTTCTTCCATTTCTACGAACTTTCAAAATTCACAC 745
 QY 614 ATCAAGAGCGTGAAGTGGGAGAGATCTGAAATGTTGGGAGAAAGAGGGGATCTC 673
 DB 746 ATCAGGCTGCCGATGGGTTAATCCCATCCACCTGTTAGGGTTAAGTGAATGGCCAGG 805
 QY 674 CAGATGATCAGAAACATTTGATAGAGAAATTTTATAGAGAAATGTTGCAAAAGAGA 733
 DB 806 AATGGGAATTTAATCCAAATCAGAAAGTATTTCAATACGTAGAGTGGGCAAGAGAAA 865
 QY 734 ACAAAAGACCAATTTACATAGTGTGAGTTGGCTTACAGAAAAGCTGACCTTGATCAA 793
 DB 866 ATACAGTACCAATCTTTCTGTGTAATTCGATGTTTTCGTAAGAGCATGAGACACAA 925
 QY 794 GAATTAATGATGATCTTTGTTGCTTGCAGAAATGGAAGAAAGAGTGGAGCTGGGAT 853
 DB 926 GGGTTAGTGTGACCAAGAGTGTGAGAAATGCGGAGAGATTTGGATTTTCAATACCGCT 965
 QY 854 ACTGGGAATTTTGTCCGGTTTGTGTATGATGATCTGAGAAAAACCTGGAATTAAG 913

Key Location/Qualifiers
 CDS 69..1484
 /tag= a
 /note= "Claim 5"

MO9814597-A1.
 09-APR-1998.
 03-OCT-1997; 97WO-US18008.
 04-OCT-1996; 96US-0027883.
 (UYGE-) UNIV GEORGIA RES FOUND INC.
 Chen H, Li X, Ljungdahl LG;
 MPI; 1998-240096/21.
 P-PSDB; AAM56742.

New recombinant DNA encoding Orpinomyces cellulase protein - useful for, e.g. producing recombinant Orpinomyces cellulase in host cell

Claim 5; Page 38-40; 69pp; English.

This cDNA clone includes a claimed coding region for cellulase celB (see AAM56742) of the anaerobic bovine rumen fungus Orpinomyces sp. PC-2. It was obtained by screening a PC-2 cDNA library for clones active on remazol brilliant blue-carboxymethylcellulose. The encoded cellulase has endoglucanase, but not cellobiohydrolase, activity. CelA and celC genes (see AAV29472-73), also obtained from Orpinomyces sp. PC-2, encode cellulases having both activities (see AAM56738-39). Recombinant DNA molecules encoding Orpinomyces cellulase proteins are claimed, as well as recombinant cells selected from Saccharomyces cerevisiae, Escherichia coli, Aspergillus, Trichoderma reesei, Pichia, Penicillium, Streptomyces or Bacillus, and a method for producing recombinant cellulase by culturing these host cells.

Sequence 1826 BP; 654 A; 236 C; 313 G; 623 T; 0 other:

Query Match 7.7%; Score 73; DB 19; Length 1826;
 Best Local Similarity 50.4%; Pred. No. 2.1e-09;
 Matches 212; Conservative 0; Mismatches 200; Indels 9; Gaps 1;

107 ATGAGTCTTCGACATTATAAAGACCGGTTCTCTCATGTCGATTCGATTAAGAT 166
 277 AAGAGCTTACTATTAACCTAGTGTGTTTCAACACTTTCCTACTACTT 336
 167 GGAGTACGACGCTTACGCGTTCTCTTAATAATCATGATCGCTCTTCAAAAGAG 226
 337 GGAGTGTGCTATTTGGTGTATGCTCTGACATATAATTTAGTGTGATGATGAAGAG 396
 227 TGGATGAAGTATAAGGAGCCCTTAAGAGAGGACTGCTGTTGCTATTAATATTCATC 286
 397 TTCATGAAGTGTGCTTATTAACACTGCTGATGATGATGATGATGATGATGATGATG 456
 287 ACTACG-----AGGAGTGAATGAATGATGATGATGATGATGATGATGATGATG 337
 457 ATGAAGCTTGAATTTATGCTTCCAAAGAAATTAAGAGATGATGATGATGATGATG 516
 338 CTTCTTGAAGAAATTTGCTGATGCTTATTAAGACTATCCGAAATCTATTTTGAAG 397
 517 CCATCTGGAAGAAATTTGCTGATGCTTATTAAGACTATTTATTTTGAAG 576
 398 TTCTGATTAACCTACGGAATCTTACCGGAAATGAATGAATGAATGAATGAATGAATGA 457
 577 GAATGAATGAATGAATGAATGAATGAATGAATGAATGAATGAATGAATGAATGAATGA 636
 458 CTTCTAAAGTATTAAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 517
 637 GTTGAATTTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 696

Qy 518 G 518
 Db 697 G 697

RESULT 8
 AAV37413
 ID AAV37413 standard; cDNA to mRNA; 1826 BP.
 XX
 AC AAV37413;
 XX
 DT 14-SEP-1998 (first entry)
 XX
 DE Orpinomyces cellulase CelB cDNA.
 XX
 KW Cellulase; endoglucanase; CelB gene; ds.
 XX
 OS Orpinomyces sp. strain PC-2.
 XX
 FH Key Location/Qualifiers
 FT CDS 69..1484
 FT /tag= a
 FT /note= "Claim 5"

MO9814597-A1.
 09-APR-1998.
 03-OCT-1997; 97WO-US18008.
 04-OCT-1996; 96US-0027883.
 (UYGE-) UNIV GEORGIA RES FOUND INC.
 Chen H, Li X, Ljungdahl LG;
 MPI; 1998-240096/21.
 P-PSDB; AAM56742.

New recombinant DNA encoding Orpinomyces cellulase protein - useful for, e.g. producing recombinant Orpinomyces cellulase in host cell

Claim 5; Page 38-40; 69pp; English.

This cDNA clone includes a claimed coding region for cellulase celB (see AAM56742) of the anaerobic bovine rumen fungus Orpinomyces sp. PC-2. It was obtained by screening a PC-2 cDNA library for clones active on remazol brilliant blue-carboxymethylcellulose. The encoded cellulase has endoglucanase, but not cellobiohydrolase, activity. CelA and celC genes (see AAV29472-73), also obtained from Orpinomyces sp. PC-2, encode cellulases having both activities (see AAM56738-39). Recombinant DNA molecules encoding Orpinomyces cellulase proteins are claimed, as well as recombinant cells selected from Saccharomyces cerevisiae, Escherichia coli, Aspergillus, Trichoderma reesei, Pichia, Penicillium, Streptomyces or Bacillus, and a method for producing recombinant cellulase by culturing these host cells.

Sequence 1826 BP; 654 A; 236 C; 313 G; 623 T; 0 other:

Query Match 7.7%; Score 73; DB 19; Length 1826;
 Best Local Similarity 50.4%; Pred. No. 2.1e-09;
 Matches 212; Conservative 0; Mismatches 200; Indels 9; Gaps 1;

107 ATGAGTCTTCGACATTATAAAGACCGGTTCTCTCATGTCGATTCGATTAAGAT 166
 277 AAGAGCTTACTATTAACCTAGTGTGTTTCAACACTTTCCTACTACTT 336
 167 GGAGTACGACGCTTACGCGTTCTCTTAATAATCATGATCGCTCTTCAAAAGAG 226
 337 GGAGTGTGCTATTTGGTGTATGCTCTGACATATAATTTAGTGTGATGATGAAGAG 396
 227 TGGATGAAGTATAAGGAGCCCTTAAGAGAGGACTGCTGTTGCTATTAATATTCATC 286

```

Db      397  TTCTGTAAGTGTTCCTTATCTGCTTACACTGGTGTATGCACTTTAAACATTCACC 456
OY      287  ACTACG-----AGGAGTTAATGATGATCCGAAGAACACAGAAAGATTTCTG 337
Db      457  ATGAACTTGAATATTCCTTCCAAAGAAATTTAGAGATGCCAAAAGATCTTAGTTG 516
OY      338  CTCCTTGGAAACAATTCGTATGCTGTATTAAGACTATCCGAAATCTATTTTGGAA 397
Db      517  CCACTGGAACAAATTCCTGCTGATTTGGATGATGACATTTAATTTTTCGAAG 576
OY      398  TTCTGAATGAACCTCAGCAATCTTACTCCGGAATAATGAAATGACTGCTTGAGAG 457
Db      577  GAATGAATGAACCAAGAAAGTTGGGATCCAGCTGAATGGACAGGTGATCAAGAG 636
OY      458  CTCCTAAGTATATAGATGATGACAAAGACACTATATTTATAGCAGCTGAAT 517
Db      637  GTTGAATTTGCTCAATGAATGAATGCCCTTTTCTTAAACTATTCGTGCACCTGGAG 696
OY      518  G 518
Db      697  G 697

RESULT 9
AAA64503
ID      AAA64503 standard; cDNA; 1532 BP.
XX
AC      AAA64503;
XX
DT      02-JAN-2001 (first entry)
XX
DE      cDNA encoding a CelE cellulase polypeptide.
XX
KW      CelE cellulase; endoglucanase; carboxymethylcellulose; beta-glucan;
        lichenin; parantitrophenyl-beta-cellobioside; cellobiose; ss.
XX
OS      Orpinomyces sp.
XX
FH      Key      Location/Qualifiers
FT      CDS      38..1472
                /tag= a
                /product= "CelE cellulase"
XX
XX      US6110720-A.
XX
XX      29-AUG-2000.
XX
XX      17-JUL-1998; 98US-0118324.
XX
XX      17-JUL-1998; 98US-0118324.
XX
XX      (UYGE-) UNIV.GEORGIA RES FOUND INC.
XX
XX      LI X, Chen H, Ljungdahl LG;
XX
XX      WPI: 2000-586355/55.
XX
XX      P-PSDB; AAB08774.
XX
XX      Novel nucleic acids encoding cellulase from Orpinomyces species PC-2
XX      PT      which produces cellobiose from carboxymethylcellulose, cellotetraose,
XX      PT      beta-glucan and lichenin
XX
XX      Claim 3; Column 9-12; 14pp; English.
XX
XX      The present sequence encodes a CelE cellulase. The polypeptide has
XX      CC      endoglucanase activity, and hydrolyses carboxymethylcellulose,
XX      CC      beta-glucan, lichenin and parantitrophenyl-beta-cellobioside, to
XX      CC      produce cellobiose. The polypeptide is derived from Orpinomyces sp.
XX      CC      PC-2, and shows homology to several other endoglucanases. The CelE
XX      CC      polypeptide has a catalytic domain to glycosyl hydrolases of Family 5
XX      CC      anaerobic bacteria. The CelE gene is devoid of introns.

```

```

SQ      Sequence 1532 BP; 494 A; 239 C; 294 G; 505 T; 0 other;
Query Match      7.2%; Score 68.6; DB 21; Length 1532;
Best Local Similarity 53.2%; Pred. No. 3e+08;
Matches 174; Conservative 0; Mismatches 144; Indels 9; Gaps 1;

OY      125  TAAAGAAGCCGGTTCCTCCTCATGTTCCAAATTAAGATGAGTACGACGCTTACG 184
Db      268  TAAAGATCAAGGTTTAAATGTTTCCTGATTCACACCTTGGACCGGCTTGGTA 327
OY      185  CGTTCTCTCTTAAATCAATGATGCTCTTCAAAAGAGTGAATGATGAATACG 244
Db      328  ATGCGACGATTAACAATAATGATGTTGGATGAGAGAGATTCATGAAGTTGTGATT 387
OY      245  GAGCCCTGAAAAGAGACTGCGCTTTCCTAATATATTCATCTCAGAGAG----- 297
Db      388  ATGCTCTTAAACAGTGAAGCTAGCTCATCTTAATATATTCACATGAAATTTGAATTATG 447
OY      298  -TTAATGATGATCCAGAACACAGCAAGAAAGATTTCTTCTTGGAAACAAATTG 355
Db      448  CTTTCTTAACAATTTACAAAGCAAAACCAATTTTASCTGATCTGGAAACAAATTG 507
OY      356  CTGATCGTTATTAAGACTATCCGAAACTTATTTTGAATTTGTAATGAACTCAGC 415
Db      508  CTGCTGAATTCGCTAATGATGATGAACATTTAATTTTGAAGGATGATGAACCAAGAA 567
OY      416  GAATCTTACTCCGAAATGGAATG 442
Db      568  AGTTGATCATCCAAATGATGATG 594

```

```

RESULT 10
AA055036
ID      AA055036 standard; cDNA; 4098 BP.
XX
AC      AA055036;
XX
DT      21-JUL-1994 (first entry)
XX
DE      Sequence of celd cDNA in clone PCNP4.
XX
KW      Cellulase; celd; PCNP4; anaerobic rumen; ss.
XX
OS      Neocallimastix patriciarum.
XX
FH      Key      Location/Qualifiers
FT      CDS      1..2400
                /tag= a
                /product= domains I and II
XX
XX      CDS      1..111
                /tag= b
                /product= N-terminus of beta-galactosidase alpha-
                FT      peptide
                FT      /note= "derived from pbluescript SK-vector"
                FT      misc-difference 112..124
                FT      /tag= c
                FT      /note= "derived from oligo linker"
                FT      CDS      2468..4098.
                FT      /tag= d
                FT      /product= domain III
XX
XX      WO9400578-A.
XX
XX      06-JAN-1994.
XX
XX      24-JUN-1993; 93WO-AU00307.
XX
XX      24-JUN-1992; 92AU-0003096.
XX
XX      (CSIR ) COMMONWEALTH SCI & IND RES ORG.
XX
XX      Aylward JH, Gobius KS, Orpin CG, Xue GP;
XX

```

DR WPI: 1994-026214/03.
 DR P-PSDB; AAR47496, AAR49102.
 XX
 PT Cloning of cellulase clones from anaerobic rumen - by isolating
 PT mRNA from culture, converting to cDNA etc. fungi, producing
 PT enzymes useful in food processing etc., and DNA for modifying
 PT rumen or silage bacteria
 XX
 PS Claim 9; Page 39-40; 71pp; English.
 XX
 CC Clone pCNp4.1 encodes celd. It was derived as follows. N.
 CC patriciarum was used to construct a cDNA library in ZAP1 (in E.
 CC coli). Transformants were selected for enzyme activity using
 CC selective media. 11 colonies were positive, and of these 10 had the
 CC same restriction pattern, and the longest of these was designated
 CC celd (pCNp4.1)(AA055036)(AAR47496, R49102). A similar process was used
 CC to isolate the xylanase clone pNX-rac (AA055037, R47497). An enzyme
 CC composition contg. celd and xylanase is claimed.
 XX
 SQ Sequence 4098 BP; 1386 A; 598 C; 835 G; 1279 T; 0 other;
 Query Match 6.6%; Score 63.2; DB 15; Length 4098;
 Best Local Similarity 50.5%; Pred. No. 1.2e-06;
 Matches 186; Conservative 0; Mismatches 173; Indels 9; Gaps 1;
 QY 99 GATAAAAGATGAGTCTTCGCACATTATAAAGAGCGGTTCTCTCATGTTCGAATTC 158
 DB 2632 GACTACTGAAGATATGTTCAGGTTTAAATGATACCAATTTAATGTTTCGTAATTC 2691
 QY 159 AATAAGATGAGTACGACGCTTACCGGTTCTCTTATAAATCATGATCGCTTCTT 218
 DB 2692 AACTACTTGGTCTGTCACCTTCGTAAGCTCCAGATTACAGATTAAATGAATGTT 2751
 QY 219 CAAAGAGTGTGATGATGATAACGAGCCCTGAAGAGGACCTGGCTGTCTATAA 278
 DB 2752 AAGAGAGTTCATGAATTTGTTATTCATCAAGATGAGCTTTCGTTATCTTAA 2811
 QY 279 TATTCATCACT-----ACGAGAGTATGATGATCCAGAACACAGAGAAAG 329
 DB 2812 TCTTCACATGAAGATGAGAACCATGCTTCTCTGAACCTTGACACATGCCAGAAAT 2871
 QY 330 ATTCTTCTCTTTCGAACAAATTCGTGATCGTTATAAGACTATCCGAAACTTATT 389
 DB 2872 TTTAGAAGATTTGGTCTCAAAATTCGTGAAGATTATTAAGATTATGATGAACCTTAAT 2931
 QY 390 TTTTGAATTCGATGAACCTCCAGGAATTCCTACCGGAAATATGGAATGACTGCT 449
 DB 2932 TTTTGAAGATTAAACGAAACCAAGAAAGATGATCTCCAGTTGAATGGACTGTGTA 2991
 QY 450 TGAGGAAG 457
 DB 2992 TCAGGAAG 2999
 RESULT 11
 AAF83293.
 ID AAF83293 standard; cDNA: 1170 BP.
 XX
 AC AAF83293;
 XX
 DT 09-JUL-2001 (first entry)
 XX
 DE P. rhizinflata egla protein catalytic domain encoding cDNA.
 XX
 KW Cellulase; egla; fungus; polysaccharide; glycosidic linkage; enzyme;
 KW animal feed; alcohol; fermentation; catalytic domain; ss.
 XX
 OS Pirmoyces rhizinflata.
 XX
 FH Key Location/Qualifiers
 FT 1.1170
 FT CDS
 FT /tag- a
 FT /product- "egla catalytic domain"

XX
 PN US6222028-BI.
 XX
 PD 24-APR-2001.
 XX
 PF 15-OCT-1999; 99US-0419459.
 XX
 PR 15-OCT-1999; 99US-0419459.
 XX
 PA (SINT-) ACAD STINICA.
 XX
 PI Liu J, Cheng K, Tsai C, Chang C;
 XX
 DR WPI: 2001-307598/32.
 DR P-PSDB; AAB62468.
 XX
 PT New isolated nucleic acid encoding a cellulase enzyme of fungus
 PT Pirmoyces rhizinflata, for producing cellulase that hydrolyzes
 PT polysaccharides containing beta-1,3 or beta-1,4 glycosidic linkages
 XX
 PS Claim 3; Columns 17-20; 16pp; English.
 XX
 CC The invention relates to a cellulase enzyme, egla, isolated from the
 CC fungus Pirmoyces rhizinflata. The egla polypeptide hydrolyzes a
 CC polysaccharide containing a beta-1,3' or beta-1,4' glycosidic linkage.
 CC The egla polynucleotide is useful for producing cellulase enzymes that
 CC can hydrolyze the glycosidic linkages in polysaccharides such as
 CC cellulose. The enzyme is also useful in a number of industrial
 CC applications where breaking down biomass is beneficial. Cellulase is also
 CC useful as a supplement in animal feed to decrease the production of fecal
 CC waste by increasing the digestibility of the feed. Cellulase is also
 CC useful to increase the efficiency of alcoholic fermentations by
 CC converting undigestible biomass into fermentable sugars. The present
 CC sequence represents a cDNA encoding the catalytic domain of egla
 CC polypeptide.
 XX
 SQ Sequence 1170 BP; 385 A; 182 C; 247 G; 356 T; 0 other;
 Query Match 6.6%; Score 62.8; DB 22; Length 1170;
 Best Local Similarity 49.7%; Pred. No. 9.5e-07;
 Matches 194; Conservative 0; Mismatches 187; Indels 9; Gaps 1;
 QY 99 GATAAAAGATGAGTCTTCGCACATTATAAAGAGCGGTTCTCTCATGTTCGAATTC 158
 DB 228 GACTACTGAAGATATGTTCAGGTTTAAATGATGACACCAATTTAATGCTTCGTAATTC 287
 QY 159 AATAAGATGAGTACGACGCTTACCGGTTCTCTTATAAATCATGATCGCTTCTT 218
 DB 288 AACTACTTGGTCTGTCACCTTCGTAAGCTCCAGATTATAAGATTGATGATAAATGCTT 347
 QY 219 CAAAGAGTGTGATGATGATAACGAGCCCTGAAGAGGACCTGGCTGTCTATAA 278
 DB 348 AAGAGAGTTCATGAAGTTTGTATTCATCAAGAGAGGAGCATTTGTTATCTTAA 407
 QY 279 TATTCATCACTAG-----AGGAGTTAATGATGATCCAGAAACACAGAGAAAG 329
 DB 408 TCTTCATCATGAAGATGAGAACCTGATCGCTTCTCTGAACCTTGATATACACCAAGAAAT 467
 QY 330 ATTCTTCTCTTTCGAACAAATTCGTGATCGTTATAAGACTATTCGGAACCTTATT 389
 DB 468 TTTAGAAAAGATCTGGTCTCAAAATTCGTGAAGATTATTAAGATTATGATGAACCTTAAT 527
 QY 390 TTTTGAATTCGATGAACCTCCAGGAATTCCTACCGGAAATATGGAATGACTGCT 449
 DB 528 CTTCGAAAGATTAAACGAAACCAAGAAAGATGATCTCCAGTTGAATGGACTGTGTA 587
 QY 450 TGAGGAAGCTTAAGATTATAAGATCAAT 479
 DB 588 TCAGGAAGTGGGATGCTTAAATGCTAT 617
 RESULT 12
 AAF83296

ID AAF83296 standard; cDNA: 1182 BP.
 XX AAF83296;
 AC
 XX
 DT 09-JUL-2001 (first entry)
 XX
 DE P. rhizinflata egla protein encoding partial cDNA.
 XX
 KW Cellulase; egla; fungus; polysaccharide; glycosidic linkage; enzyme;
 KM animal feed; alcohol; fermentation; ss.
 XX
 OS Piromyces rhizinflata.
 XX
 FH Key. Location/Qualifiers
 FT CDS 1..1182
 FT /tag= a
 FT misc_feature 13..1182
 FT /tag= b
 FT /note= "corresponds to nucleotides 331-1499 of AAF83292"
 FT
 FN US6222028-B1.
 PN 24-APR-2001.
 XX
 PD 15-OCT-1999; 99US-0419459.
 XX
 PF 15-OCT-1999; 99US-0419459.
 XX
 PR 15-OCT-1999; 99US-0419459.
 XX
 PA (SINI-) ACAD SINICA.
 XX
 PI Liu J, Cheng K, Tsai C, Chang C;
 XX
 DR WPI: 2001-307598/32.
 DR P-PSDB: AAB62471.
 XX
 PT New isolated nucleic acid encoding a cellulase enzyme of fungus
 PT Piromyces rhizinflata, for producing cellulase that hydrolyzes
 PT polysaccharides containing beta-1,3 or beta-1,4 glycosidic linkages
 PT
 XX
 PS Disclosure: Columns 7-8; 16pp; English.
 XX
 CC The invention relates to a cellulase enzyme, egla, isolated from the
 CC fungus piromyces rhizinflata. The egla polypeptide hydrolyzes a
 CC polysaccharide containing a beta-1,3' or beta-1,4' glycosidic linkage.
 CC The egla polynucleotide is useful for producing cellulase enzymes that
 CC can hydrolyze the glycosidic linkages in polysaccharides such as
 CC cellulose. The enzyme is also useful in a number of industrial
 CC applications where breaking down biomass is beneficial. Cellulase is also
 CC useful as a supplement in animal feed to decrease the production of fecal
 CC waste by increasing the digestibility of the feed. Cellulase is also
 CC useful to increase the efficiency of alcoholic fermentations by
 CC converting undigestible biomass into fermentable sugars. The present
 CC sequence represents a partial nucleic acid sequence of GST-Egla fusion
 CC protein and encodes an egla catalytic domain downstream of the GST.
 CC
 XX
 SO Sequence 1182 BP; 388 A; 184 C; 250 G; 360 T; 0 other;
 Query Match 6.6%; Score 62.8; DB 22; Length 1182;
 Best Local Similarity 49.7%; Pred. No. 9,5e-07;
 Matches 194; Conservative 0; Mismatches 187; Indels 9; Gaps 1;
 QY 99 GATTAAGATGAGTTCCTTCGACATTATAAAGACCGGTTCTTCATGTTGCAATTCC 158
 DB 240 GACACACGACGATATGTTCAAGTTTAAATGACACCAATTTATGCTCCGATATCC 299
 QY 159 AATAAGATGAGTACGACGCTTACCGCTTCCCTTATTAATCATAGATCCCTTCT 218
 DB 300 AACTACTTGGCTGCTGATCTCGGTGAAGCTCCAGATTATTAAGTGAATAATGGTT 359
 QY 219 CAAGAAGTGGATGAAGATTAACGAGCCCTGAAAGAGGACGCTGTGGTATAA 278
 DB 360 AAAGAGAGTTCATGAAGTTGTTGATTATCCATACAGACGAGCATTTGTTATCTTAA 419

QY 279 TATTCATCACTACG-----AGGAGTTATGATGATCCAGAGACACAGAGAAG 329
 DB 420 TCTTCATCATGAGAACCTCGGATCATGCTCTCTCGAAGACTTGATACAGCCAGAAAT 479
 QY 330 ATTTCTTGCTCTTTGGAACCAATTGCTGATCGTTATTAAGACTATCCGAACTCTAT 389
 DB 480 TTTGAAAGATCTGCTGCTCAAAATTCGCTGAAGATTAAAGATTAAAGAACTTAAT 539
 QY 390 TTTGAATTTCTGAAATCACTCCAGAAATCTTACTCCGGAATAATGGAATGAACCTGT 449
 DB 540 CTTCGAAGATTAAACGACACCAAGAAAGATGATACCTCCAGTTGATGAGACTGGTCA 599
 QY 450 TGAGGAGCTCTAAAGTTATATAGATCAAT 479
 DB 600 TCAGAGAGTTGGGATGCTGTTATGCTAT 629
 RESULT 13
 ID AAF83292
 AAF83292 standard; cDNA: 2110 BP.
 XX
 AC AAF83292;
 XX
 DT 09-JUL-2001 (first entry)
 XX
 DE P. rhizinflata egla protein encoding partial cDNA.
 XX
 KW Cellulase; egla; fungus; polysaccharide; glycosidic linkage; enzyme;
 KW animal feed; alcohol; fermentation; ss.
 XX
 OS Piromyces rhizinflata.
 XX
 FH Key. Location/Qualifiers
 FT CDS 3..1751
 FT /tag= a
 FT /product= "partial egla"
 FT
 FN US6222028-B1.
 PN 24-APR-2001.
 XX
 PD 15-OCT-1999; 99US-0419459.
 XX
 PF 15-OCT-1999; 99US-0419459.
 XX
 PR 15-OCT-1999; 99US-0419459.
 XX
 PA (SINI-) ACAD SINICA.
 XX
 PI Liu J, Cheng K, Tsai C, Chang C;
 XX
 DR WPI: 2001-307598/32.
 DR P-PSDB: AAB62467.
 XX
 PT New isolated nucleic acid encoding a cellulase enzyme of fungus
 PT Piromyces rhizinflata, for producing cellulase that hydrolyzes
 PT polysaccharides containing beta-1,3 or beta-1,4 glycosidic linkages
 PT
 XX
 PS Disclosure: Columns 3-6; 16pp; English.
 XX
 CC The invention relates to a cellulase enzyme, egla, isolated from the
 CC fungus piromyces rhizinflata. The egla polypeptide hydrolyzes a
 CC polysaccharide containing a beta-1,3' or beta-1,4' glycosidic linkage.
 CC The egla polynucleotide is useful for producing cellulase enzymes that
 CC can hydrolyze the glycosidic linkages in polysaccharides such as
 CC cellulose. The enzyme is also useful in a number of industrial
 CC applications where breaking down biomass is beneficial. Cellulase is also
 CC useful as a supplement in animal feed to decrease the production of fecal
 CC waste by increasing the digestibility of the feed. Cellulase is also
 CC useful to increase the efficiency of alcoholic fermentations by
 CC converting undigestible biomass into fermentable sugars. The present
 CC sequence represents a cDNA insert in plasmid pR2301-10, the open reading
 CC frame of which encodes for a partial egla polypeptide.
 CC
 XX
 SO Sequence 2110 BP; 770 A; 265 C; 413 G; 662 T; 0 other;

GenCore version 5.1.4-p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 26, 2003, 09:20:22 ; Search time 1110 Seconds
(without alignments)
13919.364 Million cell updates/sec

Title: US-09-880-729a-1

Perfect score: 954
Sequence: 1 atgggtgtgacatccttga.....gagagatagcatgataa 954

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues
Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estrov:*
6: em_estropl:*
7: em_estro:*
8: em_hic:*
9: gb_est1:*
10: gb_est2:*
11: gb_hic:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pin:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
c 1	50.2	5.3	1101	17	CNS0039L
c 2	48.8	5.1	1101	17	CNS0182P
c 3	48.6	5.1	1101	17	CNS0039G
c 4	46.8	4.9	1101	17	CNS0006J
c 5	46.8	4.9	1101	17	CNS0106X
c 6	45.8	4.8	1204	17	CNS016E2

c 7	45.4	4.8	1101	17	CNS00LVS	AL078812 Drosophila
c 8	44.8	4.7	1031	17	CNS00CF2	AL059199 Drosophila
c 9	44	4.6	915	17	CNS015X	AL106083 Drosophila
c 10	43.8	4.6	996	17	CNS00FHH	AL071063 Drosophila
c 11	43.2	4.5	1001	17	CNS01400	AL103554 Drosophila
c 12	43	4.5	621	17	AQ236641	AQ236641 RPII11-69
c 13	43	4.5	943	17	CNS002IM	AL097720 Drosophila
c 14	43	4.5	1101	17	CNS017KE	AL108152 Drosophila
c 15	42.6	4.5	942	17	CNS04HHV	AL1290956 Tetradon
c 16	42.4	4.4	740	14	BH011716	BH011716 OGI16P21
c 17	42.4	4.4	775	17	CNS004NP	AL054294 Drosophila
c 18	42.2	4.4	438	17	AQ882597	AQ882597 HS_5407_B
c 19	42.2	4.4	1101	17	CNS00XYR	AL097005 Drosophila
c 20	42	4.4	851	17	CNS015CH	AL105275 Drosophila
c 21	42	4.4	1028	14	BQ21957	BQ21957 AGENCOURT
c 22	41.8	4.4	424	17	CNS011AO	AL100026 Drosophila
c 23	41.8	4.4	997	17	CNS005TE	AL068076 Drosophila
c 24	41.4	4.3	1101	17	CNS00L8K	AL068205 Drosophila
c 25	41.4	4.3	1141	17	CNS07206	AL425596 clone BA0
c 26	41.2	4.3	632	17	A2337849	A2337849 IM0068L10
c 27	41.2	4.3	864	17	CNS01RD2	AL157287 Anopheles
c 28	41.2	4.3	1101	17	CNS0181N	AL108773 Drosophila
c 29	41	4.3	908	17	CNS01678	AL106382 Drosophila
c 30	41	4.3	1101	17	CNS0167I	AL106896 Drosophila
c 31	40.8	4.3	401	14	BQ576938	BQ576938 PESToab1
c 32	40.8	4.3	402	14	BQ739669	BQ739669 PESToab4
c 33	40.8	4.3	471	13	B1815400	B1815400 PESToab1
c 34	40.8	4.3	552	14	BQ596856	BQ596856 PESToab2
c 35	40.8	4.3	595	14	BQ633359	BQ633359 PESToab4
c 36	40.8	4.3	601	14	BQ452215	BQ452215 PESToab9
c 37	40.6	4.3	672	17	BH605324	BH605324 BQXK91YR
c 38	40.6	4.3	914	9	AU234675	AU234675 AU234675
c 39	40.4	4.2	443	9	AU006638	AU006638 AU006638
c 40	40.4	4.2	951	17	CNS015OI	AL105780 Drosophila
c 41	40.2	4.2	928	17	CNS0102F	AL088433 Drosophila
c 42	40.2	4.2	1080	17	CNS010EP	AL069494 Drosophila
c 43	40.2	4.2	1192	17	CNS03Y36	AL265803 Tetradon
c 44	40	4.2	439	12	BF530485	BF530485 602071726
c 45	40	4.2	510	9	AJ489140	AJ489140 AJ489140

ALIGNMENTS

RESULT 1
LOCUS CNS0039L/c 1101 bp DNA linear GSS 03-JUN-1999
DEFINITION .Drosophila melanogaster genome survey sequence 77 end of BAC #
BACR08110 of RPII-98 library from Drosophila melanogaster (fruit
fly), genomic survey sequence.
ACCESSION AL063926 GI:4941783
VERSION AL063926
KEYWORDS GSS.
SOURCE Drosophila melanogaster.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 1101)
REFERENCE
AUTHORS
TITLES
JOURNALS
COMMENT
Direct Submission
Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
determination of this BAC-end sequence was carried out as part of a
collaboration with the Berkeley Drosophila Genome Project (BDGP).
The BDGP is constructing a physical map of the Drosophila
melanogaster genome using these BACs. For further information
please see <http://www.fruitfly.org> The BDGP Drosophila
melanogaster BAC library was prepared by Kazutoyo Osoegawa and
Aaron Mammosser in Pieter de Jong's laboratory in the Department of
Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
NY. The library is named RPII-98 and was constructed by partial

Ecotri digestion of Drosophila DNA provided by the BDGP from the isogenic strain y2; cn bw sp, the same strain used for the BDGP's p1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

FEATURES

source

1. 1101

/organism="Drosophila melanogaster"

/db_xref="taxon:7227"

/clone="BACR08K10"

/clone_1ib="RPI-98"

/note="end : TER3"

BASE COUNT 201 a 64 c 131 g 202 t 503 others

ORIGIN

Query Match 5.1%; Score 48.6; DB 17; Length 1101;
Best Local Similarity 15.4%; Pred. No. 0.39;
Matches 88; Conservative 253; Mismatches 225; Indels 5; Gaps 1;

363 TTTAAAGACTATCCGAAACCTATTTTGAATTCGAATGAACCTCAGCAATCT 422
1075 TTMKADDDRRMAGDADKRWMDAGTWTATWMMWMTATWTDKMMWMTAKTD 1016

423 TACTCCGAAAAATGGAATGACTGCTGAGAACTCTAAAGTTATAAGATCAATTGA 482
1015 TATWMTWRTAMRABWAGDRAGKRRDPAATDAGAGRRGRRKDKDDDDDKGG 956

483 CAAAAGCACATATATATATAGCACAGTGAATGGGGGTATATCTGCCCTGAAA 542
955 KKKKAKAKWATKMDWMDKMDKMDKADDDGAGDKDDGKDKADDDTDTGT 896

543 ACTGTCGCCCAATGGAATGGAATGGAATGGAATGGAATGGAATGGAATGGAAT 602
895 KDDDDKDDMDKADKAGTGDATWMAATDMMWMAADMMWMAADMMWMAADMMW 836

603 CGAATTTACCATCAAGAGCTGAGTGGTGAAGATCTGAGAATGTTGGGAGAAA 662
835 AMKWDAMAGARTADRDMDGAKRGARRRKRRADDDKADDDDDADATTTTWT 776

663 GTGGGATCTCCGATGATCAGAAATGATGATGATGATGATGATGATGATGATG 722
775 TTTTDDDDKMTKTDWTRMAADRTMDDDDDDDDDDDDDDDDDDDDDDDDD 716

723 GTCAAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 782
715 AADADTARDDRRRRGGADAGKGTGRRRRRRRATWDR---TDAMWADAAWTTT 661

783 CCTTGATCAAGATATAAATGAGCTCTTTCGTTCCGGAATGAGAAAGAGATG 842
660 DDTDDMDKRRRRRRGGADAGKGTGRRRRRRRATWDR---TDAMWADAAWTTT 601

843 GAGCTGCACTGCGAATTTTTCGTTTGGTGTATGATCTGAGAAAAC 902
600 DAKKADRDMAKAWARARARARARARARARARARARARARARARARARARAR 541

903 CTGGAATTAAGATCTTTGAAGCTTAATA 933
540 TATWTTTWTWTTTWTWTTTWTWTTTWTWTTTWTWTTTWTWTTTWTWTTTWT 510

RESULT 4
CNS0006J 1101 bp DNA linear GSS 03-JUN-1999

LOCUS
DEFINITION
Drosophila melanogaster genome survey sequence TET3 end of BAC #

ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

AL062049
GI:4938511
Drosophila melanogaster.
Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;

Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 1101)

REFERENCE

AUTHORS
TITLE
JOURNAL

COMMENT

source

1. 1101

/organism="Drosophila melanogaster"

/db_xref="taxon:7227"

/clone="BACR01M22"

/clone_1ib="RPI-98"

/note="end : TER3"

BASE COUNT 333 a 162 c 148 g 177 t 281 others

ORIGIN

Query Match 4.9%; Score 46.8; DB 17; Length 1101;
Best Local Similarity 30.2%; Pred. No. 1;
Matches 150; Conservative 116; Mismatches 213; Indels 17; Gaps 2;

304 AATGATCCAGACACGAGCAAGAAAGATTTCTGCTTTGGAACAATGCTGATCGT 363
481 AADGAAAAAAR 540

364 TATTAAGACTATCCGAAACCTATTTTGAATTCGAATGAACCTCAGCAATCTT 423
541 GAAAAAAMWATATKAATTTTTRRTTRAGTATTTTDRAKAAAAAATAATATT 600

424 ACTCCGAAAAATGGAATGATGATGATGATGATGATGATGATGATGATGATG 483
601 AGAAAAAGAAAAAAGAAAGTGTGGAAGAAAAAADAATTAAKAKAKAAAA 660

484 AAAAGCACACTATATTAAGGCTCTTTCGTTCCGGAATGAGAAAGAGATG 528
661 GGGAAAAAATTAATAAARARARARARARARARARARARARARARARARAR 720

529 TCTGCCCTGAAAAAGCTGCTGCCAAATGGAAGAAAAAATTTCTATGATTCAC 588
721 RGRGGGRRARAAAAAKTKRKKTKTKRRAAGRRARAAAAAARAAAAADATGK 780

589 TACTACATCTCTTCCGATTTACCAAGAGCTGAGTGGTGA--GATCTGAGA 646
781 TWAARWRTAAKMDKATATAAARARARARARARARARARARARARARARAR 840

647 AATGGTTGGAAGAAAGTGGGATCTCCAGATTCGAGAAACATTTGATGAAGATCA 706
841 RRRGRRKDDTARRRAGAGDGDAAKAAWMAAATTTARWDTATADDRARAAWKRA 900

707 ATTTTATAGAAGATGCTCAAAAAGAAAGAACCAATTTACATAGTGTGATGTTG 766
901 AAADRDKDKGRABGCTGRRRRARARARARARARARARARARARARARARAR 960

767 CTTACAGAAAAGCTGA 782
961 WDDAAMAAAKAGTG 976

RESULT 5	CNS0106X	1101 bp	DNA	linear	GSS 26-JUL-1999
LOCUS	CNS0106X				
DEFINITION	Drosophila melanogaster genome survey sequence 17 end of BAC BACN03K20 of DrosBAC library from Drosophila melanogaster (fruit fly), genomic survey sequence.				
ACCESSION	AL098595				
VERSION	AL098595.1				
KEYWORDS	GSS.				
SOURCE	Drosophila melanogaster.				
ORGANISM	Drosophila melanogaster Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.				
REFERENCE	Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr)				
AUTHORS	- Web : www.genoscope.cns.fr				
TITLE	Determination of this BAC-end and sequence was carried out as part of a collaboration with the European Drosophila Genome Project (EDGP) http://www.edgp.ebi.ac.uk - This Drosophila melanogaster BAC library (Dros BAC) was made by Alain Billaud at CEPH (Centre d'Etude du Polymorphisme Humain) with funding provided by a MRC project grant. The DNA was prepared from embryos by Alain Bucheton and Genevieve Payan. It has been constructed in the vector pBeloBAC11.				
COMMENT					
FEATURES	Location/Qualifiers				
source	1..1101 /organism="Drosophila melanogaster" /db_xref="taxon:7227" /clone="BACN03K20" /clone_1ib="DrosBAC" /plasmid="pBeloBAC11" /note="end : 17"				
BASE COUNT	258 a 107 c 60 g 175 t 501 others				
ORIGIN					
Query Match	4.9%; Score 46.8; DB 17; Length 1101;				
Best Local Similarity	17.0%; Pred. No. 1;				
Matches	74; Conservative 191; Mismatches 170; Indels 1; Gaps 1				
Oy	516 ATGGGGGGTATATCTGCCCTTGAAAAACGCTGCTGCCAAAATGGGAAAAAATCTTAT 575 : : : : : : : : : : : : : : : : : : : : : : :				
Db	654 AKATKKKARDDKDTAKAMWMDKATKKAKKAKAKAKAKAKATATGDKARAKAKAK 713 : : : : : : : : : : : : : : : : : : : : : : : : : :				
Oy	576 AGTTACATTCACCTACTACATCCCTTTCGAAATTATTCACCATCAAGAGAGCTAGTGGTGA 635 : : : : : : : : : : : : : : : : : : : : : : : : : :				
Db	714 ATTKRKAARAMDITATWTATADAAADKCAKXKAKAKAKADARDKRRMDKKRRAKAKA 773 : : : : : : : : : : : : : : : : : : : : : : : : : :				
Oy	636 AGGATCTGAGA-AATGCTTGGGAAGAAGTGGGATCTCCAGATGATCAAGAAACATTGA 694 : : : : : : : : : : : : : : : : : : : : : : : : : :				
Db	774 AKAADDDADDAKADKADAKADAKKADAAADADGREGGKKRRAKDRKKKKKKDAMGDKKKAK 833 : : : : : : : : : : : : : : : : : : : : : : : : : :				
Oy	695 TAGAAGATTCATTTATATAGACAATGGTGCACAAAGAACAAAGACAATTTACATAG 754 : : : : : : : : : : : : : : : : : : : : : : : : : :				
Db	834 KDRKAAAKAKADADDAKAKRRRRADDKRKAADAKAAKKAKKDDDDKAATKAKKATKAK 893 : : : : : : : : : : : : : : : : : : : : : : : : : :				
Oy	755 GTGAGTTTGTGCTCTACAGAAAAGCTGACCTTGATCAAGATATAAATGGACCTCCTTG 814 : : : : : : : : : : : : : : : : : : : : : : : : : :				
Db	894 KDKAKAKKKKKKKKDKAKAKAKADADDKDDDKDDDKDKADKKKWKDRKADKAKK 953 : : : : : : : : : : : : : : : : : : : : : : : : : :				
Oy	815 TCGTTGGGAATGGAGAAAGAGATGGAGCTGGCTACTACTGGGAATTTGTTCCGGTT 874 : : : : : : : : : : : : : : : : : : : : : : : : : :				
Db	954 DDDKKDKKDDKDDKDDADAKAMKAKKADADAKKAKAKAKAKAADDADAADAKAADA 1013 : : : : : : : : : : : : : : : : : : : : : : : : : :				
Oy	875 TTGGTGTATTATATCTGTGAGAAAAACCTTGAATTAAGATCTTTAGAACCTTTAATG 934 : : : : : : : : : : : : : : : : : : : : : : : : : :				
Db	1014 KAKKAKDDAAKAKAKADAKADAKADAKAKAARAKAKADADAAKAKADKDDKDDAKDD 1073 : : : : : : : : : : : : : : : : : : : : : : : : : :				
Oy	935 GAGGAGATGACATTGA 950				

RESULT 6	CNS016E2/c	1204 bp	DNA	linear	GSS 26-JUL-1999
LOCUS	CNS016E2				
DEFINITION	Drosophila melanogaster genome survey sequence 17 end of BAC BACN15A12 of DrosBAC library from Drosophila melanogaster (fruit fly), genomic survey sequence.				
ACCESSION	AL106628				
VERSION	AL106628.1	GI:5622852			
KEYWORDS	GSS.				
SOURCE	Drosophila melanogaster.				
ORGANISM	Drosophila melanogaster. Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.				
REFERENCE	Genoscope. Direct Submission Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : sequef@genoscope.cns.fr - Web : www.genoscope.cns.fr) determination of this BAC-end sequence was carried out as part of a collaboration with the European Drosophila Genome Project (EDGP) - http://www.edgp.ebi.ac.uk - This Drosophila melanogaster BAC library (Dros BAC) was made by Alain Billard at CEPH (centre d'Etude du Polymorphisme Humain) with funding provided by a MRC project grant. The DNA was prepared from embryos by Alain Bucheton and Genevieve Payan. It has been constructed in the vector pBelobAC11.				
FEATURES	location/Qualifiers				
source	1..1204 /organism="Drosophila melanogaster" /db_xref="taxon:7227" /clone="BACN15A12" /clone_11b="DrosBAC" /plasmid="pBelobAC11" /note="end : 77"				
BASE COUNT	298 a 172 c 106 g 316 t 312 others				
ORIGIN					
Query Match	4.8%; Score 45.8; DB 17; Length 1204;				
Best Local Similarity	21.7%; Pred. No. 1.8; 119; Indels 0; Gaps 0;				
Matches 64; Conservative 112; Mismatches					
658	AGAAAGTGGGCGATCCACATGATGACGAACATTTGATAGAAGAAATTCATTTTATAGAA 717				
1180	ATTAATKMKKAKKAKKAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 1121				
718	GAATGCTCAAAAAAGACAAAGACCAATTTTCATAGTGGAGTGGTGCCATACAGAA 777				
1120	ADAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 1061				
778	GCAGACCTTGATCAAGCAATTAATATGACCTCTTGTGCTTGCAGGAATGAGAAAAG 837				
1060	AKKKKKKAKKADKADKADKADKADKADKADKADKADKADKADKADKADKADKADKADKAD 1001				
838	AGATGAGAGCTGGGCATACCTGGGAATTTGTCCGGTTTGGTGTTATGATCTCTGAGA 897				
1000	AAAKKDKKAKKDKKDKKDKKDKKDKKDKKDKKDKKDKKDKKDKKDKKDKKDKKDKKDK 941				
898	AAAACTCGAATTAAGATCTTTAGAGCTTAAATGAGAGAGATAGCATTTGAAT 952				
940	AAAAAAKKDKKADAKAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 886				
RESULT 7	CNS001V5/c	1101 bp	DNA	linear	GSS 14-JUN-1999
LOCUS	CNS001V5				
DEFINITION	Drosophila melanogaster genome survey sequence TE13 end of BAC: BACR8119 of RCT-98 library from Drosophila melanogaster (fruit				

ACCESSION		fily), genomic survey sequence.
VERSION	AL078812	
KEYWORDS	AL078812.1 GI:5102102	
SOURCE	GSS:	Drosophila melanogaster.
ORGANISM		Drosophila melanogaster. Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachyceta; Muscomorpha; Epiphytoidae; Drosophilidae; Drosophila. 1 (bases 1 to 1101)
REFERENCE AUTHORS TITLE JOURNAL	Submitted (11-JUN-1999) Genoscope. BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr Web : www.genoscope.cns.fr) Determination of this BACP-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACS. For further information please see http://www.fruitfly.org The BDGP Drosophila melanogaster BAC library was prepared by Kazuo Osoegawa and Aaron Mamoser at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y2; cn bw sp, the same strain used for the BDGP's p1 and Est libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPac Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.	
FEATURES		
Source	/organism="Drosophila melanogaster" Location/Qualifiers 1..1101 /db_xref="taxon:7227" /clone="BACR48119" /clone.lib="RPCTI-98" /note="end : TET3"	
BACE COUNT	206 a 185 c 183 g 354 t 173 others	
ORIGIN		
Query Match	Best Local Similarity 4.8%; Score 45.4; DB 17; Length 1101;	
Matches	Conservative 94; Conservative 47; Mismatches 106; Indels 3; Gaps 1;	
OY	TTTATAGAGANTGCTAATAAGAACAACCATTATACATAGGTGAAGTTGGTC	767
Db	KTWAAACCGMAAGCAGCKTAFAAAAAARGBRAANAACKRTAMTKAGAKGHKKMGATT	940
OY	CTACGAAAAG--CTGACTTGATCAAGATTAAGAAGCCCTGCCTTGTCTGCCGA	824
Db	AKAMGGBTAAKTWTTTAATTKAAMWTBAMTTAKMRGAGCTGARGGGGARGGARG	880
OY	AATGAGAAAAGAGCATGGAGCTGGGCATCTAGGAATTTTTGCTCCGTTTTGGCTTTA	884
Db	GGRGGRGRAAAAGRARAAAAGCDTRCAMWRARKKGCGGRAMAGCAAARAAGWTGA	820
OY	TGATCTCTGAGAAAACCTGSAATARAAGATCTTTTAGAGCTTTAATTAAGAGGAGATAG	944
Db	WAAGGAKTWRATKAWATGCGCAATADAAATWATATAAATTKKAASAMPTAA	760
OY	CATTGAATPA 954	
Db	AKTNAAASAM 750	
RESULT 8		
CNSOOCF2		
LOCUS	CNSOOCF2	
DEFINITION	Drosophila melanogaster genome survey sequence TEU3 end of BAC # fly), genomic survey sequence.	
ACCESSION	AL059199	
VERSION	GI:4946662	

KEYWORDS	GSS.
SOURCE	Drosophila melanogaster.
ORGANISM	Drosophila melanogaster Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila. 1 (bases 1 to 1031)
REFERENCE	Genoscope. Direct Submission Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : BP 191 91006 Evry cedex - FRANCE (E-mail : sequef@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT	Determination of this BAC end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see http://www.fruitfly.org The BDGP Drosophila melanogaster BAC library was prepared by Kazuhiro Osoegawa and Aaron Mammosser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPc1-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain Y2: cn bw sp, the same strain used for the BDGP's P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm . Location/Qualifiers 1..1031
FEATURES	source /organism="Drosophila melanogaster" /db_xref="taxon:7227" /clone="BACR25K05" /clone_1lb="RPc1-98" /note="end = TET3"
BASE COUNT	176 a 169 c 124 g 214 t 348 others
ORIGIN	
Query Match	4.7%; Score 44.8; DB 17; Length 1031;
Best Local Similarity	26.6%; Pred. No. 3.1;
Matches 71; Conservative 80; Mismatches 116; Indels 0; Gaps 0;	
Oy	619 GGAGCTGAGTGGGTGGAGAGATCTGAGAAATGCTTGGGAGAAAGTGGGATCTCCAGAT 678
Db	722 RGGGARGRRKKDGRARRRAGTGMWGGDKAGAGTGWKGAAMAKRTGGWKRGRKARRRW 781
Oy	679 GATCGAACAACATTTATAGAACATTCATTTATAGAAAGTGGTCAAAAAGAACAA 738
Db	782 KRTTGWDRGDTWAKGKTGRRAKKNWTRWAKRAAADADMAADTTAGAGGDMWAAAMR 841
Oy	739 AGACCAATTTACATAGTGAAGTTGGTGCCTACAGAAAAGCTGACCTTGAATCAAGAATA 798
Db	842 RGGKAGATADRTAGAMRDKRRWDKAGTDRKRTADTKRADARADADNRGAGTAWGAA 901
Oy	799 AATGAGACCTCTTGTGCTGCGCAATGGAGAAAGAGATGGAGCTGGGCATATCG 858
Db	902 WAMAWKAARTAGTGDTRKATADTADGARNRMMWATARRKAWGAGAKATWTTWTTD 961
Oy	859 GAATTTGTTCGGGTTTGGTGTAT 885
Db	962 TWTKTTTGTNTKTKTTTWTWTTTTKT 988
RESULT 9	
CNS015YX/c	
LOCUS	CNS015YX 915 bp DNA linear GSS 26-JUL-1999
DEFINITION	Drosophila melanogaster genome survey sequence Sfe end of BAC BACN15N15 of DrosBAC library from Drosophila melanogaster (fruit fly), genomic survey sequence.
ACCESSION	AL106083
VERSION	AI106083.1 GI:5619993
KEYWORDS	GSS.
SOURCE	Drosophila melanogaster. Drosophila melanogaster.
ORGANISM	


```

/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone="BACN37F11"
/clone_lib="DrosBAC"

```

```
/plasmid="pBelobAC11"
/note="end : T7"
```

	/note - end : 17		
BASE COUNT	302 a	83 c	32 g 379 t 305 others
ORIGIN			

Query Match	4.58;	Score 43;	DB 17;	length 1101;
Best Local Similarity	29.18;	Pred. No. 8.3;		
Matches 127;	Conservative 89;	Mismatches 221;	Indels 0;	Gaps 0;

234 AGTGATAACGGAGCCCTGGAAGAAGAGACTGGCTGTGCTATAATAATATTCATCACTACGA 293

Db 814 TAAAKKSKGKKGKCMKCAAWKDGAAAKKKKSCCAAKAAAAKRAWWAGKKKAKKAAGG 755

294 GGAGTTATGATGATCCAGAAGACACAGAAGATTTCTTGCTCTTTGGAAACAAT 353

Db 754 KGAKKGRGKAAAAAAAAAAKAKKKKTTAAKGTGGGAAACGKKKRAMAGDAAKKGAAAAAAAAA 695

[illegible]

694 AA AHDAADAAAAAGWAAAKKAKKKKKMAGRAKAKACAAAKKAACGTCAGACAC 63

474 ATCAATTGCAAAAAGCACACTTAATTATAGGCACAGCTGAATGGGGGGTATATCTGC 53

```

      : || : | : : || | : | : || : |||
Db 574 KWWAAAAKKKAGAGADKKKAAAGAAKKAAATATWMSGATTGKCAGKATAAGAWK 511

```

QY 534 CCTGAAAACCTGTCGTCCCAAAATGGGAAAAAATTCATATAGTTACAATTCACCTACTA 59

Db 514 AAAAAAAAAAGKAAKWAATCKAAKAAKAAKAAKATGAAATATATTAAATGTTTTGTTA 45

QY 594 CATTCCTTCGATTTA 610
||| ||| ||| |||

Db 454 TTATTTTTATTTA 438

RESULT 15

LOCUS	942 bp	DNA	linear	GSS 21-MA
CNS04HHV				
CNS04HHV				

110D1 of library G from *Tetradodon nigroviridis*, genomic survey sequence.

ACCESSION	AL290956	GI:8029536
VERSION	AL290956.1	

KEYWORDS GSS; genome survey sequence.
SOURCE *Tetrahodon nigroviridis*.

ORGANISM
Tetraodon nigroviridis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostei
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei

ACTinopterygii; Neopterygii; Teleostei; Zosterostei; Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes; Tetraodontidae; Tetraodon.

REFERENCE
1 (bases 1 to 942)
Roest-Croollius, H., Jallion, O., Dasilva, C., Bouneau, L., Fisher

Bernot, A., Fizames, C., Wincker, P., Brothier, P., Queller, F., Saurin, W. and Weissensbach, J.

TITLE
Human gene number estimate provided by genome wide analysis
Tetradon nigroviridis DNA sequence

JOURNAL Unpublished
REFERENCE 2 (bases 1 to 942)
Post-Crollius H. Tallon, O. Dasilva, C. F. Fiammes, C. Fisher

AUTHORS
ROSE CLOUTIER, M. Sc., Université de la Colombie-Britannique, Vancouver, Canada
BOUNEAU, L., Billault, A., Quetier, F., Saurin, W., Bernot, A. and Weissenbach, J.

TITLE Characterization and repeat analysis of the compact genome of

JOURNAL freshwater pufferfish Tetraodon nigroviridis
 REFERENCE Unpublished
 3 (bases 1 to 942)
 AUTHORS Genoscope.
 TITLE Direct Submission
 JOURNAL Submitted (12-APR-2000)

COMMENT This sequence is a single read and was generated as part of a large
 scale clone-end sequencing project of the Tetraodon nigroviridis
 genome. For more information, please take a look at
<http://www.genoscope.cns.fr/Tetraodon>.

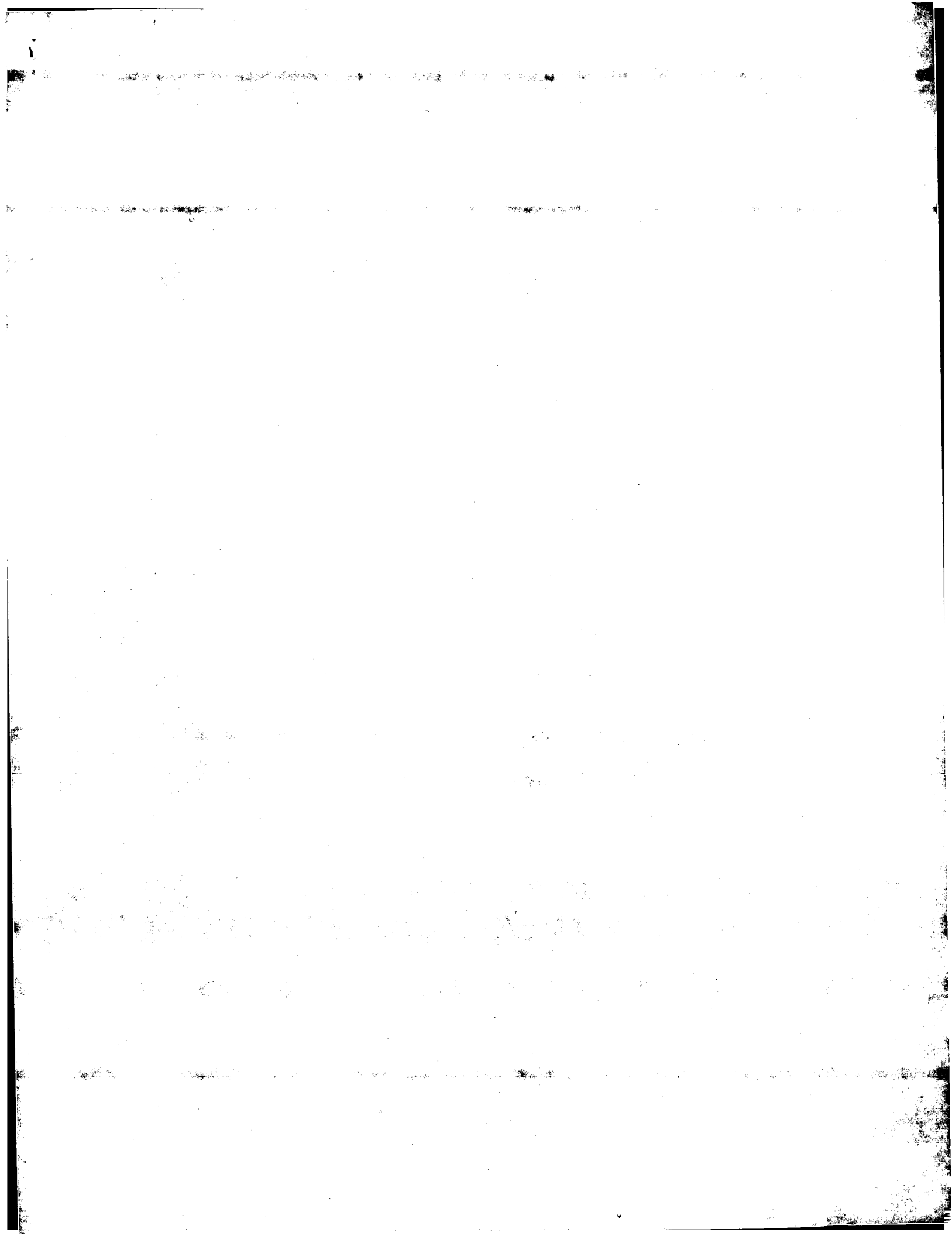
FEATURES
 source
 1.942
 /organism="Tetraodon nigroviridis"
 /db_xref="taxon:99883"
 /clone_1ib="G"
 /note="Genoscope sequence ID : C0BG110CB06SPl-end :
 PUC-Or1"

BASE COUNT 450 a 75 c 186 g 153 t 78 others
 ORIGIN

Query Match 4.5%; Score 42.6; DB 17; Length 942;
 Best Local Similarity 47.0%; Pred. No. 10;
 Matches 94; Conservative 14; Mismatches 89; Indels 3; Gaps 1;

QY 555 AAATGGGAAATAATCTTACTGATCAATTCACATCCTTCGATTTACCCA 614
 ||||| ||||| :||| | | | :||| | | | :||| : : :
 Db 336 AAATTTATGAAAAAATMTAAWAAAAAATAAAAAAATWGAAT--MTM 392
 QY 615 TCAGGAGCTGAGTGGGTGAAGGATCTGAGAAATGGTTGGGAAGAAGTGGGGATCTCC 674
 || | | | | | | | | | | | | | | | | | | | | |
 Db 393 AAAAAAARAAAAAGGTTAAAWAAAAAAGAGTAGTNNANNAANATAAAAANNAN 452
 QY 675 AGATGATCAGAACATTTGATAGAGATTCATTTTATAGAAATGTCAAAAAGAA 734
 | | | | | | | | | | | | | | | | | | | | | : : : : :
 Db 453 AAANNNTAAATAAAATTAAGAAAAAATAAAATAAAAGWAAAAAARAWAAAW 512
 QY 735 CAAAGACCAATTTACATAG 754
 || || | | | | | | | | | | | | | | | | | | | | |
 Db 513 AAATAGMAARTAAAGAAAG 532

Search completed: April 26, 2003, 10:18:56
 Job time : 1120 secs



GenCore version 5.1.4-p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 26, 2003, 09:27:47 ; Search time 46 seconds
(without alignments)
6360.209 Million cell updates/sec

Title: US-09-880-729A-1

Perfect score: 954
Sequence: 1 atgggtgttgatccttga.....gaggagatagcatgtaataa 954

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 441362 segs, 153338381 residues

Total number of hits satisfying chosen parameters: 883724

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

Issued_Patents_NA:*
1: /cgn2_6/ptodata/1/ina/5A.COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B.COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A.COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B.COMB.seq:*
5: /cgn2_6/ptodata/1/ina/PCITUS.COMB.seq:*
6: /cgn2_6/ptodata/1/ina/Backfile1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match length	ID	Description
1	954	100.0	954 2 US-09-066-075-1	Sequence 1, Appl
2	954	100.0	954 2 US-08-518-615A-1	Sequence 1, Appl
3	954	100.0	954 3 US-08-951-889-1	Sequence 1, Appl
4	954	100.0	954 4 US-09-472-857-1	Sequence 1, Appl
5	235.8	24.7	1083 4 US-09-134-078-13	Sequence 13, Appl
6	73	7.7	1826 4 US-09-286-691-11	Sequence 11, Appl
7	73	7.7	1826 4 US-09-687-147-11	Sequence 11, Appl
8	68.6	7.2	1532 3 US-09-118-324-1	Sequence 1, Appl
9	62.8	6.6	1170 4 US-09-419-459-3	Sequence 3, Appl
10	62.8	6.6	1182 4 US-09-419-459-9	Sequence 9, Appl
11	62.8	6.6	2110 4 US-09-419-459-1	Sequence 1, Appl
12	60.4	6.3	7218 4 US-08-232-463-14	Sequence 14, Appl
13	49.4	5.2	1614 4 US-09-110-959A-3	Sequence 3, Appl
14	48.2	5.1	2774 3 US-08-732-433-2	Sequence 2, Appl
15	40	4.2	2930 1 US-08-356-354-5	Sequence 5, Appl
16	40	4.2	2930 2 US-08-778-656-5	Sequence 5, Appl
17	40	4.2	3625 2 US-08-356-354-3	Sequence 3, Appl
18	40	4.2	3625 2 US-08-778-656-3	Sequence 3, Appl
19	40	4.2	3740 1 US-08-356-354-1	Sequence 1, Appl
20	40	4.2	3740 2 US-08-778-656-1	Sequence 1, Appl
21	36.4	3.8	1299 1 US-08-463-262A-1	Sequence 1, Appl
22	36.4	3.8	1299 1 US-08-463-989-1	Sequence 1, Appl
23	36.4	3.8	1299 4 US-09-003-574-1	Sequence 1, Appl
24	36.4	3.8	1299 4 US-09-003-570-1	Sequence 1, Appl
25	36.4	3.8	2028 4 US-09-003-574-33	Sequence 33, Appl
26	36.4	3.8	2028 4 US-09-003-570-33	Sequence 33, Appl
27	36.4	3.8	2032 4 US-09-003-574-32	Sequence 32, Appl

28	36.4	3.8	2032 4 US-09-003-570-32	Sequence 32, Appl
29	36.4	3.8	2076 4 US-09-003-574-30	Sequence 30, Appl
30	36.4	3.8	2076 4 US-09-003-570-30	Sequence 30, Appl
31	36.4	3.8	2126 1 US-08-463-262A-2	Sequence 2, Appl
32	36.4	3.8	2126 1 US-08-463-989-2	Sequence 2, Appl
33	36.4	3.8	2126 4 US-09-003-574-2	Sequence 2, Appl
34	36.4	3.8	2126 4 US-09-003-570-2	Sequence 2, Appl
35	36.4	3.8	2292 4 US-09-003-574-29	Sequence 29, Appl
36	36.4	3.8	2292 4 US-09-003-570-29	Sequence 29, Appl
37	35.4	3.7	2323 2 US-07-862-588B-3	Sequence 3, Appl
38	35	3.7	1074 3 US-09-248-335-67	Sequence 67, Appl
39	34.8	3.6	87350 3 US-08-781-891-79	Sequence 79, Appl
40	34.8	3.6	87543 4 US-09-791-211-3	Sequence 3, Appl
41	34.6	3.6	1816 4 US-09-149-476-167	Sequence 167, App
42	34.2	3.6	2107 4 US-09-180-852-1	Sequence 1, Appl
43	34.2	3.6	19124 2 US-08-487-826B-13	Sequence 13, Appl
44	34	3.6	2236 4 US-09-697-367-1	Sequence 1, Appl
45	34	3.6	4450 3 US-08-617-860B-2	Sequence 2, Appl

ALIGNMENTS

RESULT 1
US-09-066-075-1
; Sequence 1, Application US/09066075
; Patent No. 5925749
; GENERAL INFORMATION:
; APPLICANT: Mathur, E., et al.
; TITLE OF INVENTION: Carboxymethyl Cellulase from Thermotoga Maritima
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: CARELLI, BYRNE, BAIN, GILFILLAN,
; ADDRESS: CECCHI, STEWART & OLSTEIN
; STREET: 6 BECKER FARM ROAD
; CITY: ROSELAND
; STATE: NEW JERSEY
; COUNTRY: USA
; ZIP: 07068
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 INCH DISKETTE
; COMPUTER: IBM PS/2
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: WORD PERFECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/066,075
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/518,615
; FILING DATE: August 23, 1995
; ATTORNEY/AGENT INFORMATION:
; NAME: FERRARO, GREGORY D.
; REGISTRATION NUMBER: 36,134
; REFERENCE/DOCKET NUMBER: 331400-20
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-994-1700
; TELEFAX: 201-994-1744
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 954 BASE PAIRS
; TYPE: NUCLEIC ACID
; STRANDEDNESS: SINGLE
; TOPOLOGY: LINEAR
; MOLECULE TYPE: CDNA
; US-09-066-075-1
Query Match 100.0%; Score 954; DB 2; Length 954;
Best Local Similarity 100.0%; Pred. No. 5.4e+259;
Matches 954; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
1 ATGGGTGTGATCCTTTTAAAGAACAAATATTGGGAGGCGCTTATATAGGAAT 60
|||||

```

Db      1 ATGGGTGTGATCCTTTTGAAGAACAAATATTGGAGAGGCAATTAATAGAAAT 60
QY      61 GCGCTTGAAGCACCANAATGAGGAGACTGGGAGTGGTGAATAAAGATGCTTGGAC 120
        |||||||
Db      61 GCGCTTGAAGCACCANAATGAGGAGACTGGGAGTGGTGAATAAAGATGCTTGGAC 120
QY      121 ATTATAAAGAGGCGGTTCTCATGTTTCAATTCATTAAGATGAGTACGACGCT 180
        |||||||
Db      121 ATTATAAAGAGGCGGTTCTCATGTTTCAATTCATTAAGATGAGTACGACGCT 180
QY      121 ATTATAAAGAGGCGGTTCTCATGTTTCAATTCATTAAGATGAGTACGACGCT 180
        |||||||
Db      121 ATTATAAAGAGGCGGTTCTCATGTTTCAATTCATTAAGATGAGTACGACGCT 180
QY      181 TACGCTTCTCTCTTAAATCATGATGCTTCTTCAAAAAGTGAATGAGTGA 240
        |||||||
Db      181 TACGCTTCTCTCTTAAATCATGATGCTTCTTCAAAAAGTGAATGAGTGA 240
QY      181 TACGCTTCTCTCTTAAATCATGATGCTTCTTCAAAAAGTGAATGAGTGA 240
        |||||||
Db      241 AACGAGGCCCGAAGAGAGAGTGGCTGTGCTATTAATATTCATCAGAGAGTTA 300
        |||||||
Db      241 AACGAGGCCCGAAGAGAGAGTGGCTGTGCTATTAATATTCATCAGAGAGTTA 300
QY      301 ATGAATGATCCAGAGAACACAGAGAAATTTCTGCTCTTTGGAACCAATTCGAT 360
        |||||||
Db      301 ATGAATGATCCAGAGAACACAGAGAAATTTCTGCTCTTTGGAACCAATTCGAT 360
QY      301 ATGAATGATCCAGAGAACACAGAGAAATTTCTGCTCTTTGGAACCAATTCGAT 360
        |||||||
Db      361 CGTTTAAAGACTATCCGGAACCTATTTTGAATTCGAATGAACCTCAGGAAAT 420
        |||||||
QY      361 CGTTTAAAGACTATCCGGAACCTATTTTGAATTCGAATGAACCTCAGGAAAT 420
        |||||||
Db      421 CTACTCCGGAAGAAATGGAATGAACTGCTTGAAGAGCTCTAAAGTTAATGATCAAT 480
        |||||||
QY      421 CTACTCCGGAAGAAATGGAATGAACTGCTTGAAGAGCTCTAAAGTTAATGATCAAT 480
        |||||||
Db      421 CTACTCCGGAAGAAATGGAATGAACTGCTTGAAGAGCTCTAAAGTTAATGATCAAT 480
QY      481 GACAAAAGACACATATATTATAGGACAGCTGAATGGGGGGTATCTCCCTTGA 540
        |||||||
Db      481 GACAAAAGACACATATATTATAGGACAGCTGAATGGGGGGTATCTCCCTTGA 540
QY      481 GACAAAAGACACATATATTATAGGACAGCTGAATGGGGGGTATCTCCCTTGA 540
        |||||||
Db      541 AAACGTCTGTCCCAAAATGGGAAAAAATCTATAGTTACATTCACACTACATCAATCT 600
        |||||||
QY      541 AAACGTCTGTCCCAAAATGGGAAAAAATCTATAGTTACATTCACACTACATCAATCT 600
        |||||||
Db      601 TTGCAATTTACCCATCAAGAGCTGAGTGGTGAAGAGTCTGAGAAATGGTTGGGA 660
        |||||||
QY      601 TTGCAATTTACCCATCAAGAGCTGAGTGGTGAAGAGTCTGAGAAATGGTTGGGA 660
        |||||||
Db      661 AAGTGGGATCTCCAGATGATCAGAAACATTTGATAGAGCAATTCATTTATAGAGAA 720
        |||||||
QY      661 AAGTGGGATCTCCAGATGATCAGAAACATTTGATAGAGCAATTCATTTATAGAGAA 720
        |||||||
Db      661 AAGTGGGATCTCCAGATGATCAGAAACATTTGATAGAGCAATTCATTTATAGAGAA 720
QY      721 TGCTCAAAAAGACAAAGACCAATTTACATAGCGAGTTGGTCTACAGAAAAGCT 780
        |||||||
Db      721 TGCTCAAAAAGACAAAGACCAATTTACATAGCGAGTTGGTCTACAGAAAAGCT 780
QY      721 TGCTCAAAAAGACAAAGACCAATTTACATAGCGAGTTGGTCTACAGAAAAGCT 780
        |||||||
Db      781 GACCTGAATCAAGAAATGAATGACCTCTTGTCTGCGAATGAGAAAAGAGAGA 840
        |||||||
QY      781 GACCTGAATCAAGAAATGAATGACCTCTTGTCTGCGAATGAGAAAAGAGAGA 840
        |||||||
Db      841 TGGAGCTGGCATACTGGGAATTTTGTCCGGTTTGTGTTATGATCTGAGAAA 900
        |||||||
QY      841 TGGAGCTGGCATACTGGGAATTTTGTCCGGTTTGTGTTATGATCTGAGAAA 900
        |||||||
Db      841 TGGAGCTGGCATACTGGGAATTTTGTCCGGTTTGTGTTATGATCTGAGAAA 900
QY      901 ACCCTGAATGAAGATCTTTAGAACTTTAATAGAGAGATGACATTGATAATA 954
        |||||||
Db      901 ACCCTGAATGAAGATCTTTAGAACTTTAATAGAGAGATGACATTGATAATA 954
        |||||||

```

RESULT 2

```

US-08-518-615A-1
; Sequence 1, Application US/08518615A
; Patent No. 5962238
; GENERAL INFORMATION:
; APPLICANT: Mathur, E., et al.
; TITLE OF INVENTION: Carboxymethyl Cellulase from Thermotoga Maritima
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CARELLA, BYRNE, BAIN, GILFILLAN,

```

```

; ADDRESSEE: CECCHI, STEWART & OLSTEIN
; STREET: 6 BECKER FARM ROAD
; CITY: ROSELAND
; STATE: NEW JERSEY
; COUNTRY: USA
; ZIP: 07068
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 INCH DISKETTE
; COMPUTER: IBM PS/2
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: WORD PERFECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/518,615A
; FILING DATE: August 23, 1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: FERRARO, GREGORY D.
; REGISTRATION NUMBER: 36,134
; REFERENCE/DOCKET NUMBER: 331400-20
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-994-1700
; TELEFAX: 201-994-1744
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 954 BASE PAIRS
; TYPE: NUCLEIC ACID
; STRANDEDNESS: SINGLE
; TOPOLOGY: LINEAR
; MOLECULE TYPE: CDNA
; US-08-518-615A-1
; Query Match 100.0%; Score 954; DB 2; Length 954;
; Best Local Similarity 100.0%; Pred. No. 5,4e-259;
; Matches 954; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY      1 ATGGGTGTGATCCTTTTGAAGAACAAATATTGGAGAGGCAATTAATAGAAAT 60
        |||||||
Db      1 ATGGGTGTGATCCTTTTGAAGAACAAATATTGGAGAGGCAATTAATAGAAAT 60
QY      61 GCGCTTGAAGCACCANAATGAGGAGACTGGGAGTGGTGAATAAAGATGCTTGGAC 120
        |||||||
Db      61 GCGCTTGAAGCACCANAATGAGGAGACTGGGAGTGGTGAATAAAGATGCTTGGAC 120
QY      121 ATTATAAAGAGGCGGTTCTCATGTTTCAATTCATTAAGATGAGTACGACGCT 180
        |||||||
Db      121 ATTATAAAGAGGCGGTTCTCATGTTTCAATTCATTAAGATGAGTACGACGCT 180
QY      121 ATTATAAAGAGGCGGTTCTCATGTTTCAATTCATTAAGATGAGTACGACGCT 180
        |||||||
Db      181 TACGCTTCTCTCTTAAATCATGATGCTTCTTCAAAAAGTGAATGAGTGA 240
        |||||||
QY      181 TACGCTTCTCTCTTAAATCATGATGCTTCTTCAAAAAGTGAATGAGTGA 240
        |||||||
Db      181 TACGCTTCTCTCTTAAATCATGATGCTTCTTCAAAAAGTGAATGAGTGA 240
QY      241 AACGAGGCCCGAAGAGAGAGTGGCTGTGCTATTAATATTCATCAGAGAGTTA 300
        |||||||
Db      241 AACGAGGCCCGAAGAGAGAGTGGCTGTGCTATTAATATTCATCAGAGAGTTA 300
QY      241 AACGAGGCCCGAAGAGAGAGTGGCTGTGCTATTAATATTCATCAGAGAGTTA 300
        |||||||
Db      301 ATGAATGATCCAGAGAACACAGAGAAATTTCTGCTCTTTGGAACCAATTCGAT 360
        |||||||
QY      301 ATGAATGATCCAGAGAACACAGAGAAATTTCTGCTCTTTGGAACCAATTCGAT 360
        |||||||
Db      301 ATGAATGATCCAGAGAACACAGAGAAATTTCTGCTCTTTGGAACCAATTCGAT 360
QY      361 CGTTTAAAGACTATCCGGAACCTATTTTGAATTCGAATGAACCTCAGGAAAT 420
        |||||||
Db      361 CGTTTAAAGACTATCCGGAACCTATTTTGAATTCGAATGAACCTCAGGAAAT 420
QY      361 CGTTTAAAGACTATCCGGAACCTATTTTGAATTCGAATGAACCTCAGGAAAT 420
        |||||||
Db      421 CTACTCCGGAAGAAATGGAATGAACTGCTTGAAGAGCTCTAAAGTTAATGATCAAT 480
        |||||||
QY      421 CTACTCCGGAAGAAATGGAATGAACTGCTTGAAGAGCTCTAAAGTTAATGATCAAT 480
        |||||||
Db      421 CTACTCCGGAAGAAATGGAATGAACTGCTTGAAGAGCTCTAAAGTTAATGATCAAT 480
QY      481 GACAAAAGACACATATATTATAGGACAGCTGAATGGGGGGTATCTCCCTTGA 540
        |||||||
Db      481 GACAAAAGACACATATATTATAGGACAGCTGAATGGGGGGTATCTCCCTTGA 540
        |||||||

```

OY	541	AAACGTCGTGCCAAAATGGGAAAAAAATTCATATATTCACATTCCTCCTCAATCCT	600
Db	541	AAACTGTCTGCCAAAATGGGAAAAAAATTCATATATTCACATTCCTCCTCAATCCT	600
OY	601	TTTCAATTTTACCATCAAGAGAGCTGAGCTGGGTGGGAAGATCTGAGAANTGGTTGGGAGA	660
Db	601	TTTCAATTTTACCATCAAGAGAGCTGAGCTGGGTGGGAAGATCTGAGAANTGGTTGGGAGA	660
OY	661	AAATGGGGATTCGCAGATGATCAGAAACATTTGATAGAGAATTCATTTTATAGAGAA	720
Db	661	AAATGGGGATTCGCAGATGATCAGAAACATTTGATAGAGAATTCATTTTATAGAGAA	720
OY	721	TGGTCAAAAAAGAACAAAAGACCATTACATAGTAGTAGTTGGTGGCTACAGAAAAGCT	780
Db	721	TGGTCAAAAAAGAACAAAAGACCATTACATAGTAGTAGTTGGTGGCTACAGAAAAGCT	780
OY	781	GACCTTGATCAAGAAATTAATAATGAGCTCCCTTTGCTGGCGCAATGAGAAAGAGAGA	840
Db	781	GACCTTGATCAAGAAATTAATAATGAGCTCCCTTTGCTGGCGCAATGAGAAAGAGAGA	840
OY	841	TGGAGCTGGGCACTACTGGGAATTTTGTTCGGGTTTGGTGTTATATGACTACTGAGAAA	900
Db	841	TGGAGCTGGGCACTACTGGGAATTTTGTTCGGGTTTGGTGTTATATGACTACTGAGAAA	900
OY	901	ACCTGGAATTAAGATCTTTTGAAGCTTTAATAGAGAGATGACATTTGAATTA	954
Db	901	ACCTGGAATTAAGATCTTTTGAAGCTTTAATAGAGAGATGACATTTGAATTA	954

```

1 RESULT 3
2 US-08-951-889-1
3
4 : Sequence 1, Application US/08951869
5 : Patent No. 6008032
6
7 : GENERAL INFORMATION:
8 : APPLICANT: Mathur, E., et al.
9 : TITLE OF INVENTION: Carboxymethyl Cellulase from
10 : TITLE OF INVENTION: Thermotoga Maritima
11 : NUMBER OF SEQUENCES: 4
12
13 : CORRESPONDENCE ADDRESS:
14 : ADDRESSEE: CARELLA, BRNE, BAIN, GILFILLAN,
15 : ADDRESSEE: CECCHI, STEWART & OLSTEIN
16 : STREET: 6 BECKER FARM ROAD
17 : CITY: ROSELAND
18 : STATE: NEW JERSEY
19 : COUNTRY: USA
20
21 : ZIP: 07068
22
23 : COMPUTER READABLE FORM:
24 : MEDIUM TYPE: 3.5 INCH DISKETTE
25 : COMPUTER: IBM PS/2
26 : OPERATING SYSTEM: MS-DOOS
27 : SOFTWARE: WORD PERFECT 5.1
28
29 : CURRENT APPLICATION DATA:
30 : APPLICATION NUMBER: US/08/951,889
31
32 : FILING DATE:
33 : CLASSIFICATION: 435
34 : PRIOR APPLICATION DATA:
35 : APPLICATION NUMBER: 08/518,615
36 : FILING DATE: August 23, 1995
37
38 : ATTORNEY/AGENT INFORMATION:
39 : NAME: FERRARO, GREGORY D.
40 : REGISTRATION NUMBER: 36,134
41 : REFERENCE/DOCKET NUMBER: 331400-20
42 : TELECOMMUNICATION INFORMATION:
43 : TELEPHONE: 201-994-1700
44 : TELEFAX: 201-994-1744
45
46 : INFORMATION FOR SEQ ID NO: 1:
47 : SEQUENCE CHARACTERISTICS:
48 : LENGTH: 954 BASE PAIRS
49 : TYPE: NUCLEIC ACID
50 : STRANDEDNESS: SINGLE
51 : TOPOLOGY: LINEAR
52
53 : MOLECULE TYPE: CDNA

```

US-08-951-889-1

Query Match	100.0%;	Score 954;	DB 3;	Length 954;
Best Local Similarity	100.0%;	Pred. No. 5.4e-259;		
Matches 954; Conservative	0;	Mismatches	0;	Indels 0; Gaps 0;

QY	1	ATGGGCTGTTGATCCCTTTTGAAGAACAATAATTTGGGAAGAGCATTAATATAGAAAT	60
Db	1	ATGGGTGTTGATCTCTTTTGAAGAACAATAATTTGGGAAGAGCATTAATATAGAAAT	60
QY	61	GCCTTTGAAGCACCAGATGAGGAGAGCTGGGGAGTGGTGTATAAAGATGAGTTCTTCGAC	120
Db	61	GCCTTTGAAGCACCAGATGAGGAGAGCTGGGGAGTGGTGTATAAAGATGAGTTCTTCGAC	120
QY	121	ATTATTAAGAAGACCGGTTTCTCTCATGTTTCGAATTCCAATTAAGATGAGTACGCACGCT	180
Db	121	ATTATTAAGAAGACCGGTTTCTCTCATGTTTCGAATTCCAATTAAGATGAGTACGCACGCT	180
QY	181	TACGGGTTTCCCTCTTATTAATAATATGATGATGCGTCTCTTCAAAGAGTGAATGAAGTAT	240
Db	181	TACGGGTTTCCCTCTTATTAATAATATGATGATGCGTCTCTTCAAAGAGTGAATGAAGTAT	240
QY	241	AACGGAGCCCTGAAGAAGAGAGCTGGCTGTGCTATAAATATTGCATCCTACGAGGAGTTA	300
Db	241	AACGGAGCCCTGAAGAAGAGAGCTGGCTGTGCTATAAATATTGCATCCTACGAGGAGTTA	300
QY	301	ATGAATGATCCAGAAGAACACAGAGAAAGTTCTTGCTCTTTGGAAACAATTGCTGAT	360
Db	301	ATGAATGATCCAGAAGAACAACAGAGAAAGTTCTTGCTCTTTGGAAACAATTTGCTGAT	360
QY	361	CGTTATTAAGACCTATCCCGGAACCTCTATTTTTTGAATTTCTGAATGAACCTCAGGAAAT	420
Db	361	CGTTATTAAGACCTATCCCGGAACCTCTATTTTTTGAATTTCTGAATGAACCTCAGGAAAT	420
QY	421	CTTACTCCGGAAAAATGGAATGAACTGCTTGAGGAGACCTCTAAAAGTTATTAAGATCAAT	480
Db	421	CTTACTCCGGAAAAATGGAATGAACTGCTTGAGGAGACCTCTAAAAGTTATTAAGATCAAT	480
QY	481	GACAAAAAGCACACTATATATTTATGCGCACAGCTGATATGGGGGGTATATCTGCCCTTGAA	540
Db	481	GACAAAAAGCACACTATATATTTATGCGCACAGCTGAAATGGGGGGGTATATCTGCCCTTGAA	540
QY	541	AAACGTGCTGCCCCAAAATGGGAAAAAATTTCTATTTCAATATCCATCCATCAATTCCT	600
Db	541	AAACGTGCTGCCCCAAAATGGGAAAAAATTTCTATTTCAATATCCATCTCAATTCCT	600
QY	601	TTTGCATTTTACCCATCAAGAAGAGCTGAGTGGGTGGGAAGATCTGAGAATAGTTGGGAGAA	660
Db	601	TTTGCATTTTACCCATCAAGAAGAGCTGAGTGGGTGGGAAGATCTGAGAATAGTTGGGAGAA	660
QY	661	AAATGGGGATCTCCAGATGATCAGAAACATTTGATAGAGAATTCATTTTATAGAGAA	720
Db	661	AAATGGGGATCTCCAGATGATCAGAAACATTTGATAGAGAATTCATTTTATAGAGAA	720
QY	721	TGSGTCAAAAAGAACAAAAGCCAAATTTACATAGCTAGTTGGTGGCTACAGAAAAAGCT	780
Db	721	TGSGTCAAAAAGAACAAAAGCCAAATTTACATAGCTAGTTGGTGGCTACAGAAAAAGCT	780
QY	781	GACCTTGAATCAAAATTAATAATGAGCACTCTCTTCTGCTTGCGAAATGAGAAAGAGAG	840
Db	781	GACCTTGAATCAAAATTAATAATGAGCACTCTCTTCTGCTTGCGAAATGAGAAAGAGAG	840
QY	841	TGAGAGCTGGGCATACCTGGGAATTTGTGTTCCGGTTTGGTGTATATGATACCTGAGAAAA	900
Db	841	TGAGAGCTGGGCATACCTGGGAATTTGTGTTCCGGTTTGGTGTATATGATACCTGAGAAAA	900
QY	901	ACCTGGAATAAAGATCTTTTATAGAGCTTTATATAGAGAGATATGCAATGAATAA 954	
Db	901	ACCTGGAATAAAGATCTTTTATAGAGCTTTATATAGAGAGATATGCAATGAATAA 954	

RESULT 4
US-09-472-857-1

RESULT 4
US-09-472-857-1

Sequence 1, Application US/09472857
Patent No. 6245547
GENERAL INFORMATION:
APPLICANT: Mathur, E., et al.
TITLE OF INVENTION: Carboxymethyl Cellulase from
TITLE OF INVENTION: Thermotoga Maritima
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: CARELLA, BYRNE, BAIN, GILFILLAN,
ADDRESSEE: CECCHI, STEWART & OLSTEIN
STREET: 6 BECKER FARM ROAD
CITY: ROSELAND
STATE: NEW JERSEY
COUNTRY: USA
ZIP: 07068
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 INCH DISKETTE
COMPUTER: IBM PS/2
OPERATING SYSTEM: MS-DOS
SOFTWARE: WORD PERFECT 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/472,857
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/951,889
FILING DATE:
APPLICATION NUMBER: 08/518,615
FILING DATE: August 23, 1995
ATTORNEY/AGENT INFORMATION:
NAME: FERRARO, GREGORY D.
REGISTRATION NUMBER: 36,134
REFERENCE/DOCKET NUMBER: 331400-20
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201-994-1700
TELEFAX: 201-994-1744
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 954 BASE PAIRS
TYPE: NUCLEIC ACID
STRANDEDNESS: SINGLE
TOPOLOGY: LINEAR
MOLECULE TYPE: CDNA
US-09-472-857-1

Query Match 100.0%; Score 954; DB 4; Length 954;
Best Local Similarity 100.0%; Pred. No. 5,4e-259;
Matches 954; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGGGTTGATCGCTTTGAAAGAACAAATATTGGAGAGAGCATTATATAGAAAT 60
DB 1 ATGGGTTGATCGCTTTGAAAGAACAAATATTGGAGAGAGCATTATATAGAAAT 60
QY 61 GCGCTTGAAAGCACCAAATGAGGAGACTGGGAGTGTGATATAAAGATGAGTTCTTGAC 120
DB 61 GCGCTTGAAAGCACCAAATGAGGAGACTGGGAGTGTGATATAAAGATGAGTTCTTGAC 120
QY 121 ATTATATAAAGAGCGGTTTCTCTCATGTCGATTCGATTCGATTCGATTCGATTCGATTC 180
DB 121 ATTATATAAAGAGCGGTTTCTCTCATGTCGATTCGATTCGATTCGATTCGATTCGATTC 180
QY 181 TACGCGTTTCTCTCTTAAATATCATGATCGCTTCTTCAAAAGAGTGGATGAAGTGATA 240
DB 181 TACGCGTTTCTCTCTTAAATATCATGATCGCTTCTTCAAAAGAGTGGATGAAGTGATA 240
QY 241 AACGAGACCCCTGAAAGAGAGAGCTGGCTGCTTAAATATTCATCTACAGAGAGTTA 300
DB 241 AACGAGACCCCTGAAAGAGAGAGCTGGCTGCTTAAATATTCATCTACAGAGAGTTA 300
QY 301 ATGAATGATCAGAAAGACACAGAAAGATTCTGCTCTTTGGAACAATTTGCTGAT 360
DB 301 ATGAATGATCAGAAAGACACAGAAAGATTCTGCTCTTTGGAACAATTTGCTGAT 360

QY 361 CATTATAAAGACTATCCCGAAACCTCTATTTTGAATTCGTGAATGACCTCAGCAAT 420
DB 361 CATTATAAAGACTATCCCGAAACCTCTATTTTGAATTCGTGAATGACCTCAGCAAT 420
QY 421 CTTCCTCCGAAAGAAATGATGAGACTCTGAGAGAGCTCTAAAGTTATAGATCAATT 480
DB 421 CTTCCTCCGAAAGAAATGATGAGACTCTGAGAGAGCTCTAAAGTTATAGATCAATT 480
QY 481 GACAAAGACACACTATATATATAGCAGACAGCTCAATAGGGGGGTATATGCTCTGAA 540
DB 481 GACAAAGACACACTATATATATAGCAGACAGCTCAATAGGGGGGTATATGCTCTGAA 540
QY 541 AAAGTCTGTCGCCAAATGAGAAATATCTATAGTTACATTCATCTACTCAATCTCT 600
DB 541 AAAGTCTGTCGCCAAATGAGAAATATCTATAGTTACATTCATCTACTCAATCTCT 600
QY 601 TTGCAATTTACCATCAAGAGAGCTGAGTGGTGAAGATCTGAGAAATGGTTGGAGAA 660
DB 601 TTGCAATTTACCATCAAGAGAGCTGAGTGGTGAAGATCTGAGAAATGGTTGGAGAA 660
QY 661 AAGTGGGATCTCCAGATGATCAGAAACATTTGATAGAAATTCATTTATAGAGAA 720
DB 661 AAGTGGGATCTCCAGATGATCAGAAACATTTGATAGAAATTCATTTATAGAGAA 720
QY 721 TCGTCAAAAGACAAAGACCAATTTACATAGGTGAGTTGCTGCTCAGAAAAGCT 780
DB 721 TCGTCAAAAGACAAAGACCAATTTACATAGGTGAGTTGCTGCTCAGAAAAGCT 780
QY 781 GACCTTGAATCAAGATATAATGACCTCTTGTGCTTCCGCAATGAGAAAGAGAGA 840
DB 781 GACCTTGAATCAAGATATAATGACCTCTTGTGCTTCCGCAATGAGAAAGAGAGA 840
QY 841 TGGAGCTGGCCTACTGGAATTTGTCGCTTGTGCTTATGATCTCTGAGAAA 900
DB 841 TGGAGCTGGCCTACTGGAATTTGTCGCTTGTGCTTATGATCTCTGAGAAA 900
QY 901 ACCGTGAATTAAGATCTTTAGAGCTTTAATAGAGAGATAGCATTTGAATA 954
DB 901 ACCGTGAATTAAGATCTTTAGAGCTTTAATAGAGAGATAGCATTTGAATA 954

RESULT 5
US-09-134-078-13
Sequence 13, Application US/09134078
Patent No. 6368844
GENERAL INFORMATION:
APPLICANT: Byline, Edward J.
TITLE OF INVENTION: GLYCOSIDASE ENZYMES
NUMBER OF SEQUENCES: 72
CORRESPONDENCE ADDRESS:
ADDRESSEE: Gray Cary Ware & Freidenrich LLP
STREET: 4365 Executive Drive, Suite 1600
CITY: San Diego
STATE: CA
COUNTRY: USA
ZIP: 92121
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: Windows95
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/134,078
FILING DATE: 13-AUG-1998
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/949,026
FILING DATE: 10-OCT-1997
APPLICATION NUMBER: 60/056,916
FILING DATE: 06-DEC-1996
ATTORNEY/AGENT INFORMATION:
NAME: Halle, Lisa A.
REGISTRATION NUMBER: 38,347

REFERENCE/DOCKET NUMBER: 09010/024002
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 858/677-1456
 TELEFAX: 858/677-1465
 INFORMATION FOR SEQ ID NO: 13:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1083 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: double
 TOPOLOGY: linear
 MOLECULE TYPE: Genomic DNA
 FEATURE:
 NAME/KEY: Coding Sequence
 LOCATION: 1...1080
 US-09-134-078-13

Query Match 24.7%; Score 235.8; DB 4; Length 1083;
 Best Local Similarity 54.2%; Pred. No. 5.1e-57;
 Matches 502; Conservative 0; Mismatches 422; Indels 3; Gaps 1;

14 CTTTGAAGAGCAAAATTTGGAGAGGCAATTAATATAGAAATGCGCTTGAAGCAC 73
 149 CATTGATATCAACAAATGTAAGTAAGAGTAATATGGAATGCTTGAAGCTC 208
 74 CAATGAGGAGACTGGGAGTGTATATAAAGATGATCTTCAATTAATAAGAG 133
 209 CTTTGAAGAGCTTGGGAGTAAGATGATGATGATGATGATGATGATGATG 268
 134 CCGTTCTCTCTGTTTCAATTCATTAAGAGTACGACGCTTACGCTTCTC 193
 269 GGGGATTTGATCTTGAAGATTCCTATAGATGATGATGATGATGATGATG 328
 194 CTTATTAATATCATGATGATGATGATGATGATGATGATGATGATGATGAT 253
 329 CATATATATGACAGAGATTTCTCGAAGATTAACCATGTTGATGATGATG 388
 254 AAGAGAGCTGCTGTTGATTAATATTCATCAGAGAGATTAAGATGATGAT 313
 389 AGATATATTAACAGATATCATATGATGATGATGATGATGATGATGATGAT 448
 314 AGAATACAGAGAAATTTCTGCTTTGGAACAAATGCTGATGATGATGATG 373
 449 ATAAATACGCGGATTTGTTGGTGAATTTGAGACAGATTTGAAATTTCTT 508
 374 ATCCCAAGCTATTTTGAATTTCTGATGATGATGATGATGATGATGATG 433
 509 ACCGGGAAATCTGTTCTTTGAATATCAGACAGAGCTGCTGAACTTACAG 568
 434 AATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 493
 569 AATGAGAGCACTTTATCCAAAGTGTCTCAAGATTTATCAGGAGAGCAAT 628
 494 CTTATTAATTAAGCAGCTGATGATGATGATGATGATGATGATGATGAT 553
 629 TTGTCTATTAATGCTGCTCAAACTGGGACATATAGCGAGTGAAGCTTA 668
 554 CAAAATGAGGAAATTTCTAGTATCAATTCATCAATCTTTCGATTAATCC 613
 689 TCAATGAGGAGCAATCTGTTCTTCCATCAATCAATCTTTCGATTAATCC 745
 614 ATCAGAGCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 673
 746 ATCAGAGCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 805
 674 CAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 733
 806 AATGAGGAAATTAACCAATGAGATGATGATGATGATGATGATGATGATG 865
 734 ACAAGAGCAATTTATGATGATGATGATGATGATGATGATGATGATGAT 793
 866 ATACAGTACCAATCTTCTGTTGATGATGATGATGATGATGATGATGAT 925
 794 GAATTAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 853

1046 CATTGCAACAGCTGTGTTGGCAGAC 1072

RESULT 6
 US-09-286-691-11
 Sequence 11, Application US/09286691

GENERAL INFORMATION:
 APPLICANT: Li, Xin-Liang
 APPLICANT: Ljungdahl, Lars G.
 APPLICANT: Chen, Huizhong
 TITLE OF INVENTION: Cellulases and Coding Sequences
 FILE REFERENCE: 42-96
 CURRENT APPLICATION NUMBER: US/09/286,691
 EARLIER FILING DATE: 1999-04-05
 EARLIER APPLICATION NUMBER: US 60/027,883
 EARLIER FILING DATE: 1996-10-04
 EARLIER APPLICATION NUMBER: PCT US97/18008
 NUMBER OF SEQ ID NOS: 29
 SOFTWARE: Patentln Ver. 2.0
 SEQ ID NO 11
 LENGTH: 1826
 TYPE: DNA
 ORGANISM: Orpinomyces sp. PC-2
 FEATURE:
 NAME/KEY: CDS
 LOCATION: (69)..(1481)
 US-09-286-691-11

Query Match 7.7%; Score 73; DB 4; Length 1826;
 Best Local Similarity 50.4%; Pred. No. 3.9e-11;
 Matches 212; Conservative 0; Mismatches 200; Indels 9; Gaps 1;

107 ATGATGTTCTTCGACATTAATAAGAGCCGTTCTCTCATGTTGCAATTCATA 166
 277 AAGAGCTTACTTAAGTATGATGATGATGATGATGATGATGATGATGAT 336
 167 GAGTACGACGCTTACGCGTTCCTCTTATTAATCATGATGATGATGATG 226
 337 GAGTGTGCTATTTTGGTATGATGATGATGATGATGATGATGATGATG 396
 227 TGGATGATGATTAACGAGGAGCCCTGTAAGAGAGCTGCTGTTGATTA 286
 397 TTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 456
 287 ACTACG-----AGAGTTAATGATGATGATGATGATGATGATGATGAT 337
 457 ATGAATCTTGATTAATGCTTCCAAAGATTTAGAGATGCAAAAGATCTT 516
 338 CTCTTGGAAACAAATTCGATGATGATTAAGAGCTATCCGAAACTCTAT 397
 517 CCATCTGGAACAAATTCGATGATGATGATGATGATGATGATGATGAT 576
 398 TTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 457
 577 GAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 636
 458 CTCTAAAGTTATGATGATGATGATGATGATGATGATGATGATGATGAT 517
 637 GTTGAATTTGCTCAATTAATGATGATGATGATGATGATGATGATGAT 696
 518 G 518

Best Local Similarity 49.7%; Pred. No. 2.4e-08;
Matches 194; Conservative 0; Mismatches 187; Indels 9; Gaps 1;

```
QY 99 GATTAAGATGATGCTTCTGCACATTTATAAAGAGCCGGTTCTCTCATGTTGGAATTC 158
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 228 GACTACTGAAGATATGTTCAGGTTTAAATCGAACCAATTTAAATGCTTCGTAATTC 287
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 159 AATAAGATGAGTACGACGCTTACGGGTTTCTCTTATAAATCATGATGGCTTCTT 218
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 288 AACTACTGGTCTGGTACCTCGTGAAGCTCCAGATTATTAAGATTGATGAATAATGTT 347
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 219 CAAAGAGTGAATGAAGTATAACGAGCCCTGAAAAGAGAGACTGGCTGTCTATAAA 278
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 348 AAGAGAGTTCATGAAGTGTGTGATTATCCATACAGAACGAGCATTTGTATCTTAAA 407
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 279 TATTTCATCATAG-----AGAGTTAATGATGATCCAGAAACACAGAAAG 329
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 408 TCTTCATCATGAACCTGGAATCATGCTTCTGTGAACCTTGTATACAGCCAGAAAT 457
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 330 ATTTCCTGCTCTTGGAAACAAATGCTGATCGTTATAAGACTATCCGAACTCTATT 389
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 468 TTTAGAAAAGATCTGCTCTCAATTTGCTGAAGATTTAAGATTATGATGAACACTTAAT 527
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 390 TTTTGAATTTCTGAATGAACCTCAGGAAATCTTACTCCGAAAAATGGAATGAATGCT 449
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 528 CTTCGAAGATTAACGAAACCAAGAAAGATGATACTCCAGTTGAATGAGACTGTGTGA 587
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 450 TGAGGAAGCTCTAAAGTTAATGAATCAAT 479
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 588 TCAAGAAAGTTGGAGTCTGTATATGCTAT 617
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

RESULT 10

```
US-09-419-459-9
; Sequence 9, Application US/09419459
; Patent No. 6222028
; GENERAL INFORMATION:
; APPLICANT: Liu, Jin-Hao
; APPLICANT: Tsai, Cheng-Fang
; APPLICANT: Tsai, Cheng-Fang
; APPLICANT: Tsai, Cheng-Fang
; APPLICANT: Chang, Chia-Chieh
; TITLE OF INVENTION: CELLULOSE ENZYMES
; FILE REFERENCE: 08919/037001
; CURRENT APPLICATION NUMBER: US/09/419,459
; CURRENT FILING DATE: 1999-10-15
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 9
; LENGTH: 1182
; TYPE: DNA
; ORGANISM: Pityomyces rhizinflata
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1182)
US-09-419-459-9
```

Query Match
Best Local Similarity 49.7%; Pred. No. 2.4e-08;
Matches 194; Conservative 0; Mismatches 187; Indels 9; Gaps 1;

```
QY 99 GATTAAGATGATGCTTCTGCACATTTATAAAGAGCCGGTTCTCTCATGTTGGAATTC 158
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 240 GACTACTGAAGATATGTTCAGGTTTAAATCGAACCAATTTAAATGCTTCGTAATTC 299
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 159 AATAAGATGAGTACGACGCTTACGGGTTTCTCTTATAAATCATGATGGCTTCTT 218
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 300 AACTACTGGTCTGGTACCTCGTGAAGCTCCAGATTATTAAGATTGATGAATAATGTT 359
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 219 CAAAGAGTGAATGAAGTATAACGAGCCCTGAAAAGAGAGACTGGCTGTCTATAAA 278
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 360 AAGAGAGTTCATGAAGTGTGTGATTATCCATACAGAACGAGCATTTGTATCTTAAA 419
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 279 TATTTCATCATAG-----AGAGTTAATGATGATCCAGAAACACAGAAAG 329
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

```
Db 420 TCTTCATCATGAACCTGGAATCATGCTTCTGTGAACCTTGTATACAGCCAGAAAT 479
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 330 ATTTCCTGCTCTTGGAAACAAATGCTGATCGTTATAAGACTATCCGAACTCTATT 389
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 480 TTTAGAAAAGATCTGCTCTCAATTTGCTGAAGATTTAAGATTATGATGAACACTTAAT 539
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 390 TTTTGAATTTCTGAATGAACCTCAGGAAATCTTACTCCGAAAAATGGAATGAATGCT 449
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 540 CTTCGAAGATTAACGAAACCAAGAAAGATGATACTCCAGTTGAATGAGACTGTGTGA 599
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 450 TGAGGAAGCTCTAAAGTTAATGAATCAAT 479
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 600 TCAAGAAAGTTGGAGTCTGTATATGCTAT 629
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

RESULT 11

```
US-09-419-459-1
; Sequence 1, Application US/09419459
; Patent No. 6222028
; GENERAL INFORMATION:
; APPLICANT: Liu, Jin-Hao
; APPLICANT: Tsai, Cheng-Fang
; APPLICANT: Tsai, Cheng-Fang
; APPLICANT: Tsai, Cheng-Fang
; APPLICANT: Chang, Chia-Chieh
; TITLE OF INVENTION: CELLULOSE ENZYMES
; FILE REFERENCE: 08919/037001
; CURRENT APPLICATION NUMBER: US/09/419,459
; CURRENT FILING DATE: 1999-10-15
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 2110
; TYPE: DNA
; ORGANISM: Pityomyces rhizinflata
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (3)...(1749)
US-09-419-459-1
```

Query Match
Best Local Similarity 49.7%; Pred. No. 3e-08;
Matches 194; Conservative 0; Mismatches 187; Indels 9; Gaps 1;

```
QY 99 GATTAAGATGATGCTTCTGCACATTTATAAAGAGCCGGTTCTCTCATGTTGGAATTC 158
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 557 GACTACTGAAGATATGTTCAGGTTTAAATCGAACCAATTTAAATGCTTCGTAATTC 616
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 159 AATAAGATGAGTACGACGCTTACGGGTTTCTCTTATAAATCATGATGGCTTCTT 218
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 617 AACTACTGGTCTGGTACCTCGTGAAGCTCCAGATTATTAAGATTGATGAATAATGTT 676
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 219 CAAAGAGTGAATGAAGTATAACGAGCCCTGAAAAGAGAGACTGGCTGTCTATAAA 278
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 677 AAGAGAGTTCATGAAGTGTGTGATTATCCATACAGAACGAGCATTTGTATCTTAAA 736
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 279 TATTTCATCATAG-----AGAGTTAATGATGATCCAGAAACACAGAAAG 329
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 737 TCTTCATCATGAACCTGGAATCATGCTTCTGTGAACCTTGTATACAGCCAGAAAT 796
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 330 ATTTCCTGCTCTTGGAAACAAATGCTGATCGTTATAAGACTATCCGAACTCTATT 389
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 797 TTTAGAAAAGATCTGCTCTCAATTTGCTGAAGATTTAAGATTATGATGAACACTTAAT 856
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 390 TTTTGAATTTCTGAATGAACCTCAGGAAATCTTACTCCGAAAAATGGAATGAATGCT 449
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 857 CTTCGAAGATTAACGAAACCAAGAAAGATGATACTCCAGTTGAATGAGACTGTGTGA 916
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 450 TGAGGAAGCTCTAAAGTTAATGAATCAAT 479
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 917 TCAAGAAAGTTGGAGTCTGTATATGCTAT 946
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

```
RESULT 12
US-08-232-463-14/C
; Sequence 14, Application US/08232463
; Patent No. 5670367
GENERAL INFORMATION:
APPLICANT: DORNER, F.
APPLICANT: SCHUEFLINGER, F.
APPLICANT: FALKNER, F.G.
TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
NUMBER OF SEQUENCES: 52
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 1800 Diagonal Road, Suite 500
CITY: Alexandria
STATE: VA
COUNTRY: USA
ZIP: 22313-0299
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/232,463
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/935,313
FILING DATE:
APPLICATION NUMBER: EP 91 114 300.6
FILING DATE: 26-AUG-1991
ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 30472/114 IMMU
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703)836-9300
TELEFAX: (703)683-4109
TELEX: 899149
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 7218 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
CLONE: pTZ9pt-Fls
US-08-232-463-14

Query Match
Best Local Similarity 7.3%; Pred. No. 2,4e-07;
Matches 31; Conservative 22; Mismatches 173; Indels 0; Gaps 0;

QY 429 GGAATAATGGAATGCTGCTTGAAGAGCTTAAGATTAAAGTAAATGCAAAA 488
Db 1491 GCATCTGTAATTAATCTATCTATGCAAGTAGTTAAAGATAGAAATTTGGTACRR 1432
QY 489 GCACATATATATATAGCAGCAGCTGAAGGGGGTATATCTGCCCTGAAAAACTGTC 548
Db 1431 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1372
QY 549 TGTCCCAATATGGAATAATTTCTATAGTTACATCTACTACATCATCTTGCATTT 608
Db 1371 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1312
QY 609 TACCATCAAGAGAGCTGAGTGGTGAAGATCGAATAATGTTGGAGAAAGTGGGG 668
Db 1311 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1252
QY 669 ATCTCAGATGATCAGAAACATTTGATAGAAATTCATTTTATAGACATGCTCAA 728
Db 1251 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1192
```

```
QY 729 AAGACAAAGAACCATTTACATAGTGAGTTGCTGCTTACAGAAAGCTGACTTCA 788
Db 1191 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1132
QY 789 ATCAAGATTAATGACCTCTTGTGCTTGGGAATGAGAAAGAGATGAGCTG 848
Db 1131 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1072
QY 849 GCACATA 854
Db 1071 RRRRRR 1066

RESULT 13
US-09-110-959A-3
; Sequence 3, Application US/09110959A
; Patent No. 6268197
GENERAL INFORMATION:
APPLICANT: Schuelein, Martin
APPLICANT: Outtrup, Helle
APPLICANT: Jorgensen, Per Lina
APPLICANT: Bjornvad, Mads Eskelund
TITLE OF INVENTION: Alkaline Xyloglucanase
FILE REFERENCE: 5206-200-US
CURRENT APPLICATION NUMBER: US/09/110,959A
CURRENT FILING DATE: 1998-07-07
PRIOR APPLICATION NUMBER: 0822/97
PRIOR FILING DATE: 1997-07-07
PRIOR APPLICATION NUMBER: 1213/97
PRIOR FILING DATE: 1997-10-24
PRIOR APPLICATION NUMBER: 60/054,039
PRIOR FILING DATE: 1997-07-28
PRIOR APPLICATION NUMBER: 60/063,694
PRIOR FILING DATE: 1997-10-28
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO: 3
LENGTH: 1614
TYPE: DNA
ORGANISM: Bacillus agaradhaerens NCIMB 40482
US-09-110-959A-3

Query Match
Best Local Similarity 5.2%; Score 49.4; DB 4; Length 1614;
Matches 186; Conservative 186; Mismatches 186; Indels 15; Gaps 1;

QY 101 TAAAGATGAGTCTTTCGACATTAATAAAGACCGGTTTCTTCATGCTTGAATGCAA 160
Db 125 TAAACAGAGAGTTAAATAAAGACGATTCGATGAAGGGTATAAAGCATTCGTATCCAG 184
QY 161 TAGATGAGAGTACGACGCTTACGGGTTTCTTATATAATCATGATCGCTTCYCA 220
Db 185 TGACATGCAAAATCAATGAGGTGTTCTCCAGATTAATACGATTAATGAATATATCA 244
QY 221 AAGAGTGAATGAAGTAAAGGAGCCCTGAAGAGAGCTGGCTGTTCTTAATA 280
Db 245 ATCGGTGGAGCAAGAGATGATGAGTGGCGCTTGAGAGAGACTATATATGTTAAATG 304
QY 281 TTATATACAGAGAGGTTAATGAATGATCCAGAAACACAA-----GG 325
Db 305 TGCATCATGATCATGCTGAGTGGATGATGATGATGATGATGATGATGATGATGATG 364
QY 326 AAGATTTCTTCTTGGAAACAATTTGCTGATGATGATGATGATGATGATGATGATGATG 385
Db 365 CAAGATATACAGCTATTTGGAAACATTTGCGAAAAATTCAAAGCCACCTCCATAAGT 424
QY 386 TATTTTGAATTTGGAATGAACCTCAGGAAATTTACTCCGAAAAATGATGATGAC 445
Db 425 TGATGTTGAGAGTGCATAGAGCTGAGCTTACGAGAGAGTGGGAGAGATTCAGAAA 484
QY 446 TGCTTGAAGAGCTTAAGATTAATA 472
Db 485 ATCATCATGCTTACTTAGAGATTAA 511
```


GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 26, 2003, 10:19:02 ; Search time 99 Seconds
(without alignments)
10485.604 Million cell updates/sec

Title: US-09-880-729a-1

Perfect score: 954
Sequence: 1 atgagtgatgaccttcttga.....gagagatacatgaataca 954

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 709820 seqs, 544064369 residues

Total number of hits satisfying chosen parameters: 1419640

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : Published_Applications_NA.*
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq.*
2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq.*
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq.*
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq.*
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq.*
6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq.*
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq.*
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq.*
9: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq.*
10: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq.*
11: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq.*
12: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq.*
13: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq.*
14: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	954	100.0	954	9	US-09-880-729-1
2	235.8	24.7	1083	9	US-10-121-032-13
3	40	4.2	2930	10	US-09-376-045-5
4	40	4.2	3625	10	US-09-376-045-3
5	40	4.2	3740	10	US-09-376-045-1
6	37.8	4.0	1092	9	US-09-797-464A-10
7	37.6	3.9	11474	10	US-09-816-028A-1
8	36.8	3.9	2000	9	US-09-938-842A-4603
9	36.8	3.9	3378	10	US-09-917-800A-1443
10	36.6	3.8	2392	9	US-09-797-464A-1
11	36.4	3.8	455	9	US-09-918-995-15849
12	36.4	3.8	1168	10	US-09-822-849A-492
13	36.4	3.8	1332	10	US-09-974-300-808
14	36.4	3.8	2447	9	US-10-091-504-2093
15	36.4	3.8	2447	10	US-09-764-869-2093
16	36.4	3.8	2639	9	US-10-091-504-2092
17	36.4	3.8	2639	10	US-09-764-869-2092
18	36.4	3.8	2640	9	US-10-091-504-2091
19	36.4	3.8	2640	10	US-09-764-869-2091

20	35.8	3.8	419	9	US-10-046-935-2068	Sequence 2068, Ap
21	35.8	3.8	419	9	US-09-878-178-2068	Sequence 2068, Ap
22	35.8	3.8	419	9	US-10-146-502-2068	Sequence 2068, Ap
23	35.6	3.7	640681	10	US-09-790-988-1	Sequence 1, Appli
24	35.4	3.7	6399	10	US-09-070-927A-76	Sequence 76, Appl
25	35.2	3.7	1149	10	US-09-880-107-3374	Sequence 3374, Ap
26	35	3.7	1431	10	US-09-822-849A-517	Sequence 517, App
27	35	3.7	2204	12	US-10-044-090-768	Sequence 768, App
28	34.8	3.6	9575	9	US-10-091-504-2412	Sequence 2412, Ap
29	34.8	3.6	9575	10	US-09-764-869-2412	Sequence 2412, Ap
30	34.6	3.6	1816	9	US-09-809-391-167	Sequence 167, App
31	34.6	3.6	368004	10	US-09-949-654-3	Sequence 3, Appli
32	34.4	3.6	927	10	US-09-834-975-750	Sequence 750, App
33	34.4	3.6	927	10	US-09-834-975-751	Sequence 750, App
34	34.2	3.6	2000	9	US-09-938-842A-3728	Sequence 3728, Ap
35	34.2	3.6	2150	9	US-09-981-353-45	Sequence 45, Appl
36	34	3.6	477	9	US-09-918-995-966	Sequence 966, App
37	34	3.6	2041	10	US-09-815-242-4352	Sequence 4352, Ap
38	34	3.6	2193	10	US-09-815-242-8391	Sequence 8391, Ap
39	34	3.6	2236	10	US-09-918-909-1	Sequence 1, Appli
40	34	3.6	84539	10	US-09-962-436-36	Sequence 36, Appl
41	33.8	3.5	2000	9	US-09-938-842A-3267	Sequence 3267, Ap
42	33.8	3.5	2004	10	US-09-887-576-211	Sequence 211, App
43	33.8	3.5	2004	10	US-09-887-576-297	Sequence 297, App
44	33.8	3.5	2746	7	US-08-781-986A-336	Sequence 326, App
45	33.8	3.5	6150	9	US-10-091-504-2311	Sequence 2311, Ap

ALIGNMENTS

RESULT 1
US-09-880-729-1
Sequence 1, Application US/09880729
Publication No. US20030044956A1
GENERAL INFORMATION:
APPLICANT: Machur, E., et al.
TITLE OF INVENTION: Carboxymethyl Cellulase from Thermotoga Maritima
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESSES:
ADDRESSEE: CARELLA, BYRNE, BAIN, GILFILLAN, CECCHI, STEWART & OLSTEIN
STREET: 6 BECKER FARM ROAD
CITY: ROSELAND
STATE: NEW JERSEY
COUNTRY: USA
ZIP: 07068
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 INCH DISKETTE
COMPUTER: IBM PS/2
OPERATING SYSTEM: MS-DOS
SOFTWARE: WORD PERFECT 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/880,729
FILING DATE: 12-Jun-2001
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/09/472,857
FILING DATE: <Unknown>
APPLICATION NUMBER: 08/951,889
FILING DATE: <Unknown>
APPLICATION NUMBER: 08/518,615
FILING DATE: August 23, 1995
ATTORNEY/AGENT INFORMATION:
NAME: FERRARO, GREGORY D.
REGISTRATION NUMBER: 36,134
REFERENCE/DOCKET NUMBER: 331400-20
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201-994-1744
TELEFAX: 201-994-1744
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:

Db	329	CATTGATATTGACAGGAATTTCTCGAAGAAAGTTAACCATGTTGCGATTAGGCGCTTGG	388
Qy	254	AAAGAGACACGCTGCTTGGCTATTAATATTCATCATACAGAGAGTTAATGATGCACG	313
Db	389	AGAAATTAATTTAACAAGTAATCATCAATACGACCATTTTAAAGAACTCTATCAAGAACGG	448
Qy	314	AAGAACCAAGAGAAAGATTTCTGCTCTTTGGAAACAAATTCGTGATCTTAAAGACT	373
Db	449	ATTAATATCGCGCATGTTTGGTGGAAATTTGGAGACAGATTCGAAATTTCTTTAAAGATT	508
Qy	374	ATCCGAAACTCTAATTTTTTGAATTCGTGAATGAACTCAGCGAAATCTTACTCGGAAA	433
Db	509	ACCCGAAAAATCTGTTCTTTGAATCTACAAACGAGCTGCTCAGAACTTGACAGCTGAAA	568
Qy	434	AATGGAATGAACTGCTTGAGAGAACCTCTAAAGTTTAAATGATCAATTGCAAAAAAGCACA	493
Db	569	AATGGAACGCACTTATATCCAAAAGGTGCTCAAAAGTTTATACAGGAGACAAATCCACCGGA	628
Qy	494	CTAATATTAATAGCACAGCTGGAATGGGGGGGTATATCTGCGCTTGAAAATAGTCTGTCC	553
Db	629	TTGTCATTTATTCGATGCTCCCAACTGGGACACATATAGCCGATGAGAGATCTTAAATTTAG	688
Qy	554	CAAAATGGGAAAAAATTTCTATAGTATCAAAATTCATCTACATCAATCCTTTGCAATTTAAC	613
Db	689	TCAA---CGACAAACGCATCATGTTTCTTCATTACTAGACACCTTTCAAAATTCACAC	745
Qy	614	ATCAGAGAGCTGAGTGGGTGGAAAGATCTGAGAAATGTTGGGAGAAAGTGGGATCTC	673
Db	746	ATTCGGGTGGCGGAATGGGTTAATCCATCCACCTGTAAAGGTTAAGTGGAAATGGCGAGG	805
Qy	674	CAGATGATCAAGAAACATTTGATGAGAAGATTTCAATTTTATAGAAGATGGTCAAAAAAGA	733
Db	806	AATGGGAATTTAACCAATATAGAAATGTCATTTCAAAATACGTGAGTACTGGCGAAAGCAA	865
Qy	734	ACAAAGACCAATTTACATAGTGTGAGTTTGGTGCCTACAGAAAAGCTGACTTGAATCAA	793
Db	866	ATAACGTACCAATCTTTCTTGGAATTCGTCCTATTCAAAAGACAGACATGACTCAA	925
Qy	794	GAAATAAATGAGCCTCTCTTTCGTCGCCAATATGAGAAAAGAGATGAGACTGGCAT	853
Db	926	GGGTTAAGTGGACGGAAGATGTAGAAAATATGGCGGAAGAAATTTGGATTTTCATACGCGT	985
Qy	854	ACTGGGAATTTTGTCCGCTTTTGGTGTATGATACTCTGAGAAAACCTGGAATTAAG	913
Db	986	ATGCGGATTTTGTGCGAGATTTTGGCATATAGCATATGATGTCTCAAACTGTATCGAAC	1044
Qy	914	ATCTTTAGAACTTTAATATGACGAG	940
Db	1046	CATTGCCAACAGCTGTGTGGTGCACAG	1072
RESULT 3			
US-09-376-045-5			
Sequence 5, Application US/09376045			
Patent No. US20020019998A1			
GENERAL INFORMATION:			
APPLICANT: Aventis CropScience GmbH			
TITLE OF INVENTION: DNA SEQUENCES AND PLASMIDS FOR THE PREPARATION OF			
FILE OF INVENTION: PLANTS WITH CHANGED SUCROSE CONCENTRATION			
FILE REFERENCE: 514413-3772			
CURRENT APPLICATION NUMBER: US/09/376,045			
CURRENT FILING DATE: 1999-08-17			
NUMBER OF SEQ. ID NOS: 6			
SOFTWARE: PatentIn Ver. 2.1			
SEQ ID NO 5			
LENGTH: 2930			
TYPE: DNA			
ORGANISM: Solanum tuberosum			
FEATURE:			
NAME/KEY: CDS			
LOCATION: (118) ..(2841)			
US-09-376-045-5			

[illegible]

```

RESULT 4
US-09-376-045-3
: Sequence 3, Application US/09376045
: Patent No. US20020019998A1
: GENERAL INFORMATION:
: APPLICANT: Aventis CropScience GmbH
: TITLE OF INVENTION: DNA SEQUENCES AND PLASMIDS FOR THE PREPARATION OF
: TITLE OF INVENTION: PLANTS WITH CHANGED SUCROSE CONCENTRATION
: FILE REFERENCE: 514413-3772
: CURRENT APPLICATION NUMBER: US/09/376,045
: CURRENT FILING DATE: 1999-08-17
: NUMBER OF SEQ ID NOS: 6
: SOFTWARE: PatentIn Ver. 2.1
: SEQ ID NO 3
: LENGTH: 3625
: TYPE: DNA
: ORGANISM: Solanum tuberosum
: FEATURE:
: NAME/KEY: CDS
: LOCATION: (121)..(3282)
US-09-376-045-3

Query Match 4.2%, Score 40; DB 10; Length 3625;
Best Local Similarity 55.9%; Pred. No. 2.2;
Matches 76; Conservative 0; Mismatches 60; Indels 0; Gaps 0.

QY 225 AGTGGATGAAGTGAATAACGACGCCCTGAAAAAGAGGACTGCTGTCTATTAATTTCA 284
      || ||| || | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1881 AGCTTATGTCTCCCAATGATGACCCACAAAANAATGAGGACCTGTGATATACATAGGCT 1940

QY 285 TCATCAGCAGGAGGTAAATGAATGATCCAGAGACACACAAAGAAATTTCTCTCTTTG 344
      || ||| || | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1941 TCTTGACAAATGCTCTCTTAGTAGGATCCCATGATCAGCAGGCAATTGCTGATGCTCTTTT 2000

QY 345 GAACCAAAATTCCTCAT 360
      || | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 2001 GAAGTTGGTGTCTCAT 2016

RESULT 5
US-09-376-045-1
: Sequence 1, Application US/09376045
: Patent No. US20020019998A1
: GENERAL INFORMATION:
: APPLICANT: Aventis CropScience GmbH
: TITLE OF INVENTION: DNA SEQUENCES AND PLASMIDS FOR THE PREPARATION OF
: TITLE OF INVENTION: PLANTS WITH CHANGED SUCROSE CONCENTRATION
: FILE REFERENCE: 514413-3772
: CURRENT APPLICATION NUMBER: US/09/376,045
: CURRENT FILING DATE: 1999-08-17
: NUMBER OF SEQ ID NOS: 6
: SOFTWARE: PatentIn Ver. 2.1
: SEQ ID NO 1
: LENGTH: 3740
: TYPE: DNA
: ORGANISM: Solanum tuberosum

```

1000

778 CCTGACCTTGAATCAAGATAAAATGGACCTCCTTTGTCGTCGCGAATGGAGAAAGG 837

778 CCTGACCTTGAATCAAGATAAAATGGACCTCCTTTGTCGTCGCGAATGGAGAAAGG 837

Db 1584 AAGATCTGAGTGAATAATTAAGAAAGATCTGATGTTTATTTTGAAGG 1525
 Oy 838 AGAT 841
 Db 1524 AAT 1521

RESULT 9

US-09-917-800A-1443
 ; Sequence 1443, Application US/09917800A
 ; Patent No. US20020119462A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Mendrick, Donna
 ; APPLICANT: Porter, Mark
 ; APPLICANT: Johnson, Kory
 ; APPLICANT: Castle, Arthur
 ; APPLICANT: Elashoff, Michael
 ; APPLICANT: Gene Logic, Inc.
 ; TITLE OF INVENTION: Molecular Toxicology Modeling
 ; FILE REFERENCE: 44921-5038-US
 ; CURRENT APPLICATION NUMBER: US/09/917, 800A
 ; CURRENT FILING DATE: 2001-07-31
 ; PRIOR APPLICATION NUMBER: US 60/222,040
 ; PRIOR FILING DATE: 2000-07-31
 ; PRIOR APPLICATION NUMBER: US 60/222,880
 ; PRIOR FILING DATE: 2000-11-02
 ; PRIOR APPLICATION NUMBER: US 60/290,029
 ; PRIOR FILING DATE: 2001-05-11
 ; PRIOR APPLICATION NUMBER: US 60/290,645
 ; PRIOR FILING DATE: 2001-05-15
 ; PRIOR APPLICATION NUMBER: US 60/292,336
 ; PRIOR FILING DATE: 2001-05-22
 ; PRIOR APPLICATION NUMBER: US 60/295,798
 ; PRIOR FILING DATE: 2001-06-06
 ; PRIOR APPLICATION NUMBER: US 60/297,457
 ; PRIOR FILING DATE: 2001-06-13
 ; PRIOR APPLICATION NUMBER: US 60/298,884
 ; PRIOR FILING DATE: 2001-06-19
 ; PRIOR APPLICATION NUMBER: US 60/303,459
 ; PRIOR FILING DATE: 2001-07-09
 ; NUMBER OF SEQ ID NOS: 1740
 ; SOFTWARE: Patentln Ver. 2.1
 ; SEQ ID NO 1443
 ; LENGTH: 3378
 ; TYPE: DNA
 ; ORGANISM: Rattus norvegicus
 ; FEATURE:
 ; OTHER INFORMATION: Genbank Accession No. US20020119462A1 U01914
 US-09-917-800A-1443

Query Match 3.9%; Score 36.8; DB 10; Length 3378;
 Best Local Similarity 56.7%; Pred. No. 15;
 Matches 68; Conservative 0; Mismatches 52; Indels 0; Gaps 0;
 Oy 252 GAAAGAGAGCTGGCTGCTAATAATATCATCTACGAGAGTTAATGATGCTC 311
 Db 1542 GAAAGAGAGCTGGCTGCTAATAATATCATCTACGAGAGTTAATGATGCTC 311
 Oy 312 AGAAGAACACAGGAAGATTTCTGCTCTTTGGAACAAATGCTGATGTTATAACA 371
 Db 1602 AGAAGAACACAGGAAGATTTCTGCTCTTTGGAACAAATGCTGATGTTATAACA 371

RESULT 10

US-09-797-464A-1
 ; Sequence 1, Application US/09797464A
 ; Publication No. US20030022807A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Wiltning, Reinhard
 ; APPLICANT: Bjornvad, Mads Eskelund
 ; APPLICANT: Kauppinen, Markus Sakari
 ; APPLICANT: Schulten, Martin
 ; TITLE OF INVENTION: Family 5 Xyloglucanases

; FILE REFERENCE: 6073.200-US
 ; CURRENT APPLICATION NUMBER: US/09/797,464A
 ; CURRENT FILING DATE: 2002-02-19
 ; NUMBER OF SEQ ID NOS: 25
 ; SOFTWARE: Patentln version 3.1
 ; SEQ ID NO 1
 ; LENGTH: 2392
 ; TYPE: DNA
 ; ORGANISM: Paenibacillus pabuli
 US-09-797-464A-1

Query Match 3.8%; Score 36.6; DB 9; Length 2392;
 Best Local Similarity 49.7%; Pred. No. 14;
 Matches 93; Conservative 0; Mismatches 94; Indels 0; Gaps 0;

Oy 99 GATAAAAGATGATGTTCTTCGACATTTAATAAGAGCCGGTTCTCTCATGTTGCAATTC 158
 Db 950 GGTAACTCCAGAGTTAATCAAAAAGTAAAGCGCGCTTCAATCCATTCCTATTCC 1009
 Oy 159 AATAAGATGGAGTACGACCGCTTACCGCTTCTTATTAATATCATGATGCTCTT 218
 Db 1010 CGTTTCTATTGACACACATTTGAGACGCTCCCATTTATACATTTAATGCGCATGCT 1069
 Oy 219 CAAAAGATGATGATGATTAACGAGCCCTGAAAAGAGAGCTGCTGTTGCTATAA 278
 Db 1070 GAATCGAATTCAGCAAGTCGTGACCTATCGGTACATGAAGGCTGTATGATCATCA 1129
 Oy 279 TATTTCAT 285
 Db 1130 TATTTCAT 1136

RESULT 11

US-09-918-995-15849/c
 ; Sequence 15849, Application US/09918995
 ; Publication No. US20030073623A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Hyseq, Inc.
 ; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
 ; FILE REFERENCE: 20411-756
 ; CURRENT APPLICATION NUMBER: US/09/918,995
 ; CURRENT FILING DATE: 2001-07-30
 ; PRIOR APPLICATION NUMBER: US/09/235,076
 ; PRIOR FILING DATE: 1999-01-20
 ; NUMBER OF SEQ ID NOS: 38054
 ; SOFTWARE: FastSeq for Windows Version 3.0
 ; SEQ ID NO 15849
 ; LENGTH: 455
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 ; FEATURE:
 ; NAME/KEY: misc.feature
 ; LOCATION: (1)...(455)
 ; OTHER INFORMATION: n = A,T,C or G
 US-09-918-995-15849

Query Match 3.8%; Score 36.4; DB 9; Length 455;
 Best Local Similarity 56.8%; Pred. No. 6.2;
 Matches 67; Conservative 0; Mismatches 51; Indels 0; Gaps 0;

Oy 368 AAGACATCCCGAAGACCTATTTTGAATTCGATGAACCTCAGCAATCTTCTC 427
 Db 265 AAGTGAACACAGTTATTTGGAATTTCTTTATGAAGAAATTAAGTGAAGACATGATGTC 206
 Oy 428 CGGAAAATGGAATGAAGCTCTTGAGAGACCTTAAGAGTTATTAAGATTAATGACAA 485
 Db 205 TCGAAGAGTGAAGTAATCTGAGAGAGCAGCTCAAAAAGCTATAAATTTGTGGAGAA 148

RESULT 12

US-09-822-849A-492/c
 ; Sequence 492, Application US/09822849A

```
; Patent No. US20020045170A1
; GENERAL INFORMATION:
; APPLICANT: Wong, Gordon G.
; APPLICANT: Clark, Hilary
; APPLICANT: Fechtel, Kim
; APPLICANT: Agostino, Michael J.
; APPLICANT: Howes, Steven H.
; APPLICANT: Resnick, Richard J.
; APPLICANT: Gulukota, Kamalakhar
; APPLICANT: Graham, James R.
; APPLICANT: Genetics Institute, Inc.
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING NOVEL SECRETED PROTEINS
; FILE REFERENCE: C1N 6A03
; CURRENT APPLICATION NUMBER: US/09/822,849A
; PRIOR FILING DATE: 2001-09-04
; PRIOR APPLICATION NUMBER: 60/195,582
; PRIOR FILING DATE: 2000-04-06
; NUMBER OF SEQ ID NOS: 598
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 492
; LENGTH: 1168
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-822-849A-492
```

```
Query Match          3.8%; Score 36.4; DB 10; Length 1168;
Best Local Similarity 56.8%; Pred. No. 11;
Matches 67; Conservative 0; Mismatches 51; Indels 0; Gaps 0;
```

```
OY 368 AAGACTATCCCGAACTCTATTTTGAATTTCTGAATGTAAGTCAAGTAATCTTACTC 427
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 502 AAGTGAACAGGTATTTGGAATTTCTTATGAAGATTAAGTGAAGAACTGATAGTC 443
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 428 CGGAAATGGAATGAACTCTTGAAGAGCTCTAAAGTTAATGAATCAATGACAA 485
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 442 TGGAAAGTGAAGTAATCTGAGTAGGCACCTACAAAGATTAATTTGTGAGAA 385
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
```

```
RESULT 13
US-09-974-300-808
; Sequence 808, Application US/09974300
; Patent No. US20020146721A1
; GENERAL INFORMATION:
; APPLICANT: Berka, Randy M.
; APPLICANT: Clausen, Ib Groth
; TITLE OF INVENTION: Methods For Monitoring Multiple Gene
; FILE REFERENCE: 10085.500-US
; CURRENT APPLICATION NUMBER: US/09/974,300
; CURRENT FILING DATE: 2001-10-05
; PRIOR APPLICATION NUMBER: 09/680,598
; PRIOR FILING DATE: 2000-10-06
; PRIOR APPLICATION NUMBER: 60/279,526
; PRIOR FILING DATE: 2001-03-27
; NUMBER OF SEQ ID NOS: 8481
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 808
; LENGTH: 1332
; TYPE: DNA
; ORGANISM: Bacillus licheniformis
US-09-974-300-808
```

```
Query Match          3.8%; Score 36.4; DB 10; Length 1332;
Best Local Similarity 66.7%; Pred. No. 11;
Matches 52; Conservative 0; Mismatches 26; Indels 0; Gaps 0;
```

```
OY 341 TTGGAACAATGCTGATCTTAATAAGACTATCCGAACCTCTATTTTGAATTC 400
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 140 TCTGAGAGCAGATGCGCGCTTCAAAACAAAGTGAACGCCCTCTGTTGAATTG 199
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 401 TGAATGACCTCAGCGAA 418
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 200 TCAATGAGCCGACCGGAA 217
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
```

```
RESULT 14
US-10-091-504-2093/C
; Sequence 2093, Application US/10091504
; Publication No. US20030059908A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC007C1
; CURRENT APPLICATION NUMBER: US/10/091,504
; CURRENT FILING DATE: 2002-03-07
; NUMBER OF SEQ ID NOS: 2442
; Prior Application removed - See File Wrapper or Palm
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2093
; LENGTH: 2447
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-091-504-2093
```

```
Query Match          3.8%; Score 36.4; DB 9; Length 2447;
Best Local Similarity 56.8%; Pred. No. 16;
Matches 67; Conservative 0; Mismatches 51; Indels 0; Gaps 0;
```

```
OY 368 AAGACTATCCCGAACTCTATTTTGAATTTCTGAATGTAAGTCAAGTAATCTTACTC 427
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 495 AAGTGAACAGGTATTTGGAATTTCTTATGAAGATTAAGTGAAGAACTGATAGTC 436
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 428 CGGAAATGGAATGAACTCTTGAAGAGCTCTAAAGTTAATGAATCAATGACAA 485
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 435 TGGAAAGTGAAGTAATCTGAGTAGGCACCTACAAAGATTAATTTGTGAGAA 378
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
```

```
RESULT 15
US-09-764-869-2093/C
; Sequence 2093, Application US/09764869
; Patent No. US20020061521A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC007
; CURRENT APPLICATION NUMBER: US/09/764,869
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - refer to PALM or file wrapper
; NUMBER OF SEQ ID NOS: 2442
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2093
; LENGTH: 2447
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-869-2093
```

```
Query Match          3.8%; Score 36.4; DB 10; Length 2447;
Best Local Similarity 56.8%; Pred. No. 16;
Matches 67; Conservative 0; Mismatches 51; Indels 0; Gaps 0;
```

```
OY 368 AAGACTATCCCGAACTCTATTTTGAATTTCTGAATGTAAGTCAAGTAATCTTACTC 427
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 495 AAGTGAACAGGTATTTGGAATTTCTTATGAAGATTAAGTGAAGAACTGATAGTC 436
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 428 CGGAAATGGAATGAACTCTTGAAGAGCTCTAAAGTTAATGAATCAATGACAA 485
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 435 TGGAAAGTGAAGTAATCTGAGTAGGCACCTACAAAGATTAATTTGTGAGAA 378
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
```

Search completed: April 26, 2003, 10:58:01
Job time : 129 secs

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Comphen Ltd.

OM protein - protein search, using sw model

Run on: April 22, 2003, 13:49:58 ; Search time 40 Seconds
(without alignments)
1056.011 Million cell updates/sec

Title: US-09-880-729a-2
Perfect score: 1732
Sequence: 1 MGVDPERNKLIGRCINIGN.....LRKTWNKDLLEALLIGDSIE 317

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-Processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: /SID52/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:*
2: /SID52/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:*
3: /SID52/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:*
4: /SID52/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:*
5: /SID52/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:*
6: /SID52/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:*
7: /SID52/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:*
8: /SID52/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:*
9: /SID52/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:*
10: /SID52/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:*
11: /SID52/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:*
12: /SID52/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:*
13: /SID52/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:*
14: /SID52/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:*
15: /SID52/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:*
16: /SID52/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:*
17: /SID52/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:*
18: /SID52/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:*
19: /SID52/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:*
20: /SID52/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:*
21: /SID52/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:*
22: /SID52/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:*
23: /SID52/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1732	100.0	317	20	AAV39338 Carboxymethyl cell
2	1732	100.0	317	20	AAV23764 A carboxymethyl ce
3	1732	100.0	317	21	AAV56814 T. maritima thermo
4	1732	100.0	317	22	AAE08543 Thermotoga maritim
5	835.5	48.2	360	19	AAW49870 Thermotoga OC1/4V
6	812.5	46.9	360	18	AAW34566 Thermotoga OC1/4V
7	315	18.2	452	15	AAW49102 Translated sequenc
8	314	18.1	800	15	AAW47496 Translated sequenc
9	312	18.0	477	21	AAW08774 Amino acid sequenc
10	309	17.8	406	11	AAW08199 Neutral cellulase

11	308	17.8	390	22	AAW62468
12	308	17.8	394	22	AAW62471
13	308	17.8	582	22	AAW62467
14	306.5	17.7	574	17	AAW00383
15	298.5	17.2	329	19	AAW35002
16	296.5	17.1	471	19	AAW56742
17	292	16.9	395	22	AAE09784
18	285	16.5	400	22	AAE09785
19	283.5	16.4	348	22	AAE09788
20	281.5	16.3	537	20	AAW8463
21	279	16.1	363	22	AAE09791
22	264.5	15.3	566	12	AAW13228
23	203	11.7	160	22	AAE09789
24	203	11.7	167	22	AAE09786
25	181	10.5	410	18	AAW12378
26	181	10.5	411	18	AAW12379
27	181	10.5	411	18	AAW12381
28	181	10.5	412	18	AAW12380
29	179	10.3	409	9	AAW1843
30	177	10.2	467	17	AAW05731
31	177	10.2	467	17	AAW05731
32	170	9.8	352	20	AAW08472
33	169.5	9.8	369	20	AAW14150
34	168	9.7	400	18	AAW23601
35	168	9.7	400	18	AAW22521
36	168	9.7	400	19	AAW57431
37	168	9.7	462	19	AAW57433
38	165	9.5	472	19	AAW37243
39	163.5	9.4	499	14	AAW42122
40	163	9.4	403	23	AAW91820
41	163	9.4	405	21	AAW43848
42	159	9.2	551	18	AAW18790
43	158.5	9.2	476	19	AAW37241
44	156.5	9.0	484	19	AAW43910
45	155	8.9	386	19	AAW37242

ALIGNMENTS

RESULT 1
AAV39338
ID AAV39338 standard; Protein; 317 AA.
XX
AC AAV39338;
XX
DT 01-DEC-1999 (first entry)
XX
DE Carboxymethyl cellulase (CMCase).
XX
KW Carboxymethyl cellulase; CMCase; thermostable enzyme; hydrolysis;
KW beta 1,4 glycosidic bond; cellulose; plant biomass conversion; fuel;
KW chemical; detergent; textile industry; animal feed; waste treatment;
KW fruit juice extraction; clarification; brewing industry.
XX
OS Thermotoga maritima.
XX
PN US5962258-A.
XX
PD 05-OCT-1999.
XX
PF 23-AUG-1995; 95US-0518615.
XX
PR 23-AUG-1995; 95US-0518615.
XX
PI (DIVE-) DIVERSA CORP.
XX
PA Mathur EJ, Lam DE;
XX
DR WPI: 1999-571269/48.
XX
DR N-PSDB: AA221563.
XX
PT Recombinant polynucleotide sequence encoding a carboxymethyl cellulase

P. rhizinflata egl
P. rhizinflata egl
P. rhizinflata egl
Bacillus cellulase
Thermotoga maritim
Orpinomyces cellu
Paenibacillus pabu
Paenibacillus pabu
P. pabuli xyloglu
Bacillus agaradnae
Paenibacillus pabu
Endoglucanase enco
Paenibacillus pabu
Paenibacillus pabu
P300-CE1B fusion c
P300-CE1B fusion c
P300-CE1B fusion c
P300-CE1B fusion c
Sequence of alkali
Cellulase. Bacill
Bacillus cellulase
Actinomyces sp. 3
Protein encoded by
Bacillus agaradher
Bacillus agaradher
Bacillus agaradher
Cloned alkaline en
Globodera rostochi
NK-1 cellulase. B
Herbicideally activ
Amino acid sequenc
Corrected Bacillus
Heterodera glycine
Heterodera glycine
Globodera rostochi

PF useful for the conversion of plant biomass -
 PS Claim 1; Fig 1; 14pp: English.

CC This is the amino acid sequence of the Thermotoga maritima carboxymethyl
 CC cellulase (CMCase). CMCase is a thermostable enzyme (optimum temperature
 CC of 60 + degrees C), that catalyses the hydrolysis of the beta 1,4
 CC glycosidic bonds in cellulose. The enzyme degrades cellulose by the
 CC hydrolysis of this bond. The CMCase enzyme has a molecular weight of
 CC approximately 35kD. CMCase is used to catalyse enzymatic degradation of
 CC cellulose and can be used for the conversion of plant biomass into fuels
 CC and chemicals. CMCase can also be used in detergents, the textile
 CC industry, animal feed, waste treatment and for the clarification and
 CC extraction of juices in the fruit juice and brewing industries. The
 CC CMCase gene and 15 bp fragments of it can be used as probes to isolate
 CC nucleic acid molecules encoding CMCase enzymes from genomic libraries or
 CC to screen for related sequences. Cells expressing the CMCase
 CC polynucleotide can be used as immunogens to produce antibodies which can
 CC then be used to isolate the enzyme from cells which express CMCase. The
 CC antibodies can also be used to screen for similar enzymes from other
 CC organisms and samples.

XX Sequence 317 AA:

Query Match 100.0%; Score 1732; DB 20; Length 317;
 Best Local Similarity 100.0%; Pred. No. 3.8e-149; Indels 0; Gaps 0;
 Matches 317; Conservative 0; Mismatches 0;

QY 1 MGVDPERKKIIGRGINIGNALAPNEGDMGVVINKDEFDIIKEAGSFHVRIPIRMSTHA 60
 DB 1 MGVDPERKKIIGRGINIGNALAPNEGDMGVVINKDEFDIIKEAGSFHVRIPIRMSTHA 60
 QY 61 YAFPPYKIMDRFRRVDEVINGALKRGLAVAINIHHEELMNDPEHKEKRLALMKQIAD 120
 DB 61 YAFPPYKIMDRFRRVDEVINGALKRGLAVAINIHHEELMNDPEHKEKRLALMKQIAD 120
 QY 121 RYRDYDETLFEFLINPHGNLTPKKNELLEALKVIRSIDKKHTIIIGTAEMGGSIALE 180
 DB 121 RYRDYDETLFEFLINPHGNLTPKKNELLEALKVIRSIDKKHTIIIGTAEMGGSIALE 180
 QY 181 KLSVPKMEKNSIYTIHYNPFETHOGAEWEGSEKWLGRKWSPPDOKHLIEEFNITEE 240
 DB 181 KLSVPKMEKNSIYTIHYNPFETHOGAEWEGSEKWLGRKWSPPDOKHLIEEFNITEE 240
 QY 241 WSKKNRPRIYIGFGAYRRADLESRIKWTSPVYREMEKRRMSMAWEPSCSGFYVDTLRK 300
 DB 241 WSKKNRPRIYIGFGAYRRADLESRIKWTSPVYREMEKRRMSMAWEPSCSGFYVDTLRK 300
 QY 301 TWNKDLLEALIGDSIE 317
 DB 301 TWNKDLLEALIGDSIE 317

RESULT 2
 ID AAY23764 standard; Protein: 317 AA.

AC AAY23764;
 DT 10-SEP-1999 (first entry)
 XX A carboxymethyl cellulase enzyme.

DE Carboxymethyl cellulase enzyme; hydrolysis; cellulose; plant biomass;
 KW fuel; chemical; detergent; textile industry; animal feed;
 KW waste treatment; fruit juice; brewing industry; clarification;
 KW juice extraction.

XX Thermotoga maritima.

OS US9525749-A.

PN 20-JUL-1999.

XX 23-AUG-1995; 9505-0518615.

XX 23-AUG-1995; 9505-0518615.

PR 24-APR-1998; 9805-0066075.

XX (DIVE-) DIVERSA CORP.

XX Lam DE, Mathur EJ;

XX WPI; 1999-429058/36.

DR N-PSDB; AAX85879.

XX New carboxymethyl cellulase enzyme is useful in cellulose digestion

XX Disclosure; Fig 1A-C; 13pp: English.

XX The present sequence represents a carboxymethyl cellulase enzyme from
 CC Thermotoga maritima. Oligonucleotide probes derived from the
 CC polynucleotide sequence can be used to screen a library of genomic
 CC DNA for recovery of the polynucleotide, as PCR primers to synthesise
 CC full length polynucleotides or as diagnostic probes useful for
 CC screening for similar enzymes from other organisms and samples.
 CC The carboxymethyl cellulase enzymes can be used to catalyse hydrolysis
 CC of cellulose for the conversion of plant biomass into fuels and
 CC chemicals, for use in detergents, the textile industry, in animal feed,
 CC waste treatment and in the fruit juice/brewing industry for the
 CC clarification and extraction of juices.

XX Sequence 317 AA:

Query Match 100.0%; Score 1732; DB 20; Length 317;
 Best Local Similarity 100.0%; Pred. No. 3.8e-149; Indels 0; Gaps 0;
 Matches 317; Conservative 0; Mismatches 0;

QY 1 MGVDPERKKIIGRGINIGNALAPNEGDMGVVINKDEFDIIKEAGSFHVRIPIRMSTHA 60
 DB 1 MGVDPERKKIIGRGINIGNALAPNEGDMGVVINKDEFDIIKEAGSFHVRIPIRMSTHA 60
 QY 61 YAFPPYKIMDRFRRVDEVINGALKRGLAVAINIHHEELMNDPEHKEKRLALMKQIAD 120
 DB 61 YAFPPYKIMDRFRRVDEVINGALKRGLAVAINIHHEELMNDPEHKEKRLALMKQIAD 120
 QY 121 RYRDYDETLFEFLINPHGNLTPKKNELLEALKVIRSIDKKHTIIIGTAEMGGSIALE 180
 DB 121 RYRDYDETLFEFLINPHGNLTPKKNELLEALKVIRSIDKKHTIIIGTAEMGGSIALE 180
 QY 181 KLSVPKMEKNSIYTIHYNPFETHOGAEWEGSEKWLGRKWSPPDOKHLIEEFNITEE 240
 DB 181 KLSVPKMEKNSIYTIHYNPFETHOGAEWEGSEKWLGRKWSPPDOKHLIEEFNITEE 240
 QY 241 WSKKNRPRIYIGFGAYRRADLESRIKWTSPVYREMEKRRMSMAWEPSCSGFYVDTLRK 300
 DB 241 WSKKNRPRIYIGFGAYRRADLESRIKWTSPVYREMEKRRMSMAWEPSCSGFYVDTLRK 300
 QY 301 TWNKDLLEALIGDSIE 317
 DB 301 TWNKDLLEALIGDSIE 317

RESULT 3
 ID AAY56814 standard; Protein: 317 AA.

AC AAY56814;

DT 31-MAR-2000 (first entry)

XX T. maritima thermostable CMCase enzyme.

XX Endoglucanase; carboxymethyl cellulase; cellulose; CMCase; plant biomass;
 KW beta-1,4-glycosidic bond; detergent; textile; guar gum; animal feed;
 KW waste product; fuel; brewing; thermostable.

XX OS Thermotoga maritima.
 XX PN US6008032-A.
 XX PD 28-DEC-1999.
 XX PF 16-OCT-1997; 97US-0951889.
 XX PR 23-AUG-1995; 95US-0518615.
 XX PA (DIVE-) DIVERSA CORP.
 XX PI Mathur EJ, Lam DE;
 XX DR WPI: 2000-105335/09.
 XX DR N-PSDB; AAZ46759.
 PT Endoglucanase with carboxymethyl cellulase activity from Thermotoga
 PT maritima, useful for degrading biomass to fuels and chemicals
 XX
 PS Claim 1; Fig 1A-C; 14pp; English.
 XX CC This represents the T. maritima endoglucanase with carboxymethyl
 CC cellulase activity. The enzyme CMCase is thermostable and used to degrade
 CC carboxymethyl cellulose (CMC) or more generally any oligosaccharide that
 CC contains beta-1,4-glycosidic bonds, particularly in detergents, textiles,
 CC guar gum, animal feed, plant biomass or waste products. The CMCase enzyme
 CC is especially used to convert plant biomass to chemicals and fuels, and
 CC also for clarification and extraction of juice (e.g. in the brewing
 CC industry). CMCase can also be used to raise specific antibodies.
 XX
 SQ Sequence 317 AA;
 Query Match 100.0%; Score 1732; DB 21; Length 317;
 Best Local Similarity 100.0%; Pred. No. 3.8e-149;
 Matches 317; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MGVDPEFRNKLGRGINIGNALAPNEGDMGVKDEFDIIKAGFSHVRIPIRMSTHA 60
 DB 1 MGVDPEFRNKLGRGINIGNALAPNEGDMGVKDEFDIIKAGFSHVRIPIRMSTHA 60
 QY 61 YAFPPYKIMDRFEKRVDEVINGALKRGLAVAINIHHEELMNDPEEKERFLAMKOIAD 120
 DB 61 YAFPPYKIMDRFEKRVDEVINGALKRGLAVAINIHHEELMNDPEEKERFLAMKOIAD 120
 QY 121 RYKDYPTLPELLEINEPHGNLTPEKNWELLEALKVIRSIDKHTIIIGTAEMGGSIALE 180
 DB 121 RYKDYPTLPELLEINEPHGNLTPEKNWELLEALKVIRSIDKHTIIIGTAEMGGSIALE 180
 QY 181 KLSVPKKEKNSIVTIHYNPFEEFHOGAEWEGSEKWLGRKWSPPDOCKHLEEFNITEE 240
 DB 181 KLSVPKKEKNSIVTIHYNPFEEFHOGAEWEGSEKWLGRKWSPPDOCKHLEEFNITEE 240
 QY 241 WSKKNNRPRIYIGFAGYRKADLESRIKWTSPVYVEMEKRRMSNAWYFEGSGFGYDTLRLK 300
 DB 241 WSKKNNRPRIYIGFAGYRKADLESRIKWTSPVYVEMEKRRMSNAWYFEGSGFGYDTLRLK 300
 QY 301 TWNKDLLEALIGDSDIE 317
 DB 301 TWNKDLLEALIGDSDIE 317
 RESULT 4
 ID AAE08543 standard; Protein; 317 AA.
 AC AAE08543;
 XX
 DT 15-NOV-2001 (first entry)
 XX Thermotoga maritima CMCase enzyme.
 XX

KM Carboxymethyl cellulose; CMCase; plant biomass; animal feed; detergent;
 KM textile industry; waste treatment; brewing industry; immunogen;
 KM fruit juice industry.
 XX OS Thermotoga maritima.
 XX PN US6245547-B1.
 XX PD 12-JUN-2001.
 XX PF 27-DEC-1999; 99US-0472857.
 XX PR 23-AUG-1995; 95US-0518615.
 XX PR 16-OCT-1997; 97US-0951889.
 XX PA (DIVE-) DIVERSA CORP.
 XX PI Mathur EJ, Lam DE;
 XX DR WPI: 2001-528380/58.
 XX DR N-PSDB; AAD15243.
 PT Polynucleotide isolated from thermophile eubacteria useful for
 PT degradation of cellulose encodes an enzyme having a fully defined amino
 PT acid sequence
 XX
 PS Claim 1; Fig 1; 14pp; English.
 XX CC The present sequence is a thermostable enzyme, an
 CC endoglucanase having carboxymethyl cellulase activity referred to as
 CC (CMCase). The CMCase enzyme is useful for the degradation of cellulose,
 CC for conversion of plant biomass into fuels and chemicals, in the
 CC textile industry, in detergents, in animal feed, in waste treatment,
 CC in fruit juice/brewing industry for clarification and extraction of
 CC juices, to generate probes for identifying sequences encoding similar
 CC enzymes from other organisms, as immunogen to produce monoclonal or
 CC polyclonal antibodies.
 XX
 SQ Sequence 317 AA;
 Query Match 100.0%; Score 1732; DB 22; Length 317;
 Best Local Similarity 100.0%; Pred. No. 3.8e-149;
 Matches 317; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MGVDPEFRNKLGRGINIGNALAPNEGDMGVKDEFDIIKAGFSHVRIPIRMSTHA 60
 DB 1 MGVDPEFRNKLGRGINIGNALAPNEGDMGVKDEFDIIKAGFSHVRIPIRMSTHA 60
 QY 61 YAFPPYKIMDRFEKRVDEVINGALKRGLAVAINIHHEELMNDPEEKERFLAMKOIAD 120
 DB 61 YAFPPYKIMDRFEKRVDEVINGALKRGLAVAINIHHEELMNDPEEKERFLAMKOIAD 120
 QY 121 RYKDYPTLPELLEINEPHGNLTPEKNWELLEALKVIRSIDKHTIIIGTAEMGGSIALE 180
 DB 121 RYKDYPTLPELLEINEPHGNLTPEKNWELLEALKVIRSIDKHTIIIGTAEMGGSIALE 180
 QY 181 KLSVPKKEKNSIVTIHYNPFEEFHOGAEWEGSEKWLGRKWSPPDOCKHLEEFNITEE 240
 DB 181 KLSVPKKEKNSIVTIHYNPFEEFHOGAEWEGSEKWLGRKWSPPDOCKHLEEFNITEE 240
 QY 241 WSKKNNRPRIYIGFAGYRKADLESRIKWTSPVYVEMEKRRMSNAWYFEGSGFGYDTLRLK 300
 DB 241 WSKKNNRPRIYIGFAGYRKADLESRIKWTSPVYVEMEKRRMSNAWYFEGSGFGYDTLRLK 300
 QY 301 TWNKDLLEALIGDSDIE 317
 DB 301 TWNKDLLEALIGDSDIE 317
 RESULT 5
 ID AAW49870 standard; Protein; 360 AA.
 AC AAW49870
 XX

[illegible]

Db	231	N-DKRIIVSFHYEDPEKFTHOGAEWNPIDPPVRYKKNAGEEMLNQIRSH-----FKYVSDM	285
Oy	242	SKNKRPRYIGEEFGAYRKADLESRIKWTSEVYREMEKRSMAYWEFCGFGYDTLRKT	301
Db	286	AKONNVPIFLGEGFAYSKADMDSRVKMTESYRKMAEEFGFSYAWFCAGFGIYDRMSON	345
Oy	302	WNKDLLEALIG 312	
Db	346	WIEPLATAVVG 356	
RESULT 6			
ID	AAW34566	standard; Protein; 360 AA.	
XX	AAW34566;		
AC			
XX			
DT	12-MAR-1998	(first entry)	
XX			
DE	Thermotoga OCI/4V	endoglucanase.	
XX			
KW	Glycosidase; thermostable; textile; food processing; pharmaceutical;		
KW	detergent; baking; industry; Thermococcus; Staphylothermus;		
KW	Pyrococcus; glucose; soluble oligosaccharide; endoglucanase.		
XX			
XX	Thermotoga sp.		
PN	MO9725417-A1.		
XX			
PD	17-JUL-1997.		
XX			
PF	10-JAN-1997;	97WO-US00092.	
XX			
PR	13-SEP-1996;	96US-0712612.	
XX			
PR	11-JAN-1996;	96US-0583787.	
XX			
PA	(RECO-) RECOMBINANT BIOCATALYSIS INC.		
XX			
PI	Bylina EJ, Lam DE, Mathur EJ, Swanson RV;		
XX			
DR	WPI; 1997-372858/34.		
XX	N-PSDB; AAT93690.		
PT	New thermostable glycosidase(s) - from Thermococcus, Staphylothermus		
PT	and Pyrococcus, used in the textile, food processing,		
XX	pharmaceutical, detergent and baking industries		
PS	Claim 4; Fig 13; 82pp: English.		
XX			
CC	The present sequence represents endoglucanase isolated from Thermotoga.		
CC	The enzyme or its encoding nucleic acid sequence is used for generating		
CC	glucose from soluble oligosaccharides. The enzyme can be used in the		
CC	food processing, pharmaceutical, textile, detergent and baking		
CC	industries. The enzyme is also used to treat lactose intolerance, as a		
CC	diagnostic reporter molecule, in corn wet milling or in the fruit juice		
CC	industry. The enzymes can be used to hydrolyse guar gum to remove		
CC	non-reducing terminal mannose residues. The nucleic acids encoding the		
CC	enzyme may be used to generate probes to identify similar sequences.		
XX			
SO	Sequence	360 AA;	
Query Match			
Best Local Similarity 46.9%; Score 812.5; DB 18; Length 360;			
Matches 151; Conservative 53; Mismatches 98; Indels 9; Gaps 3;			
Oy	6	FEKNNILGRCINIGNALAEANEEDMGVVIKDEFFDIKEAGFSHVRIPIRWSTHAYAFP	65
Db	51	FEYKNMGKGVNIGNALAEAPPEGAMGVRIEHEFEIINKRGFDSVRPIRMSAHISEKPP	110
Oy	66	YKIMDRFKRKYDEVINGALKRGALVAINIHHEFLMNDPEHREKRFALQAJADYKDY	125
Db	111	YDIQNFLEPNAHVADALENNLIVIIINTHFEELYDEPKYGDVLEIYRQALAKFKKY	170

```

QY 126 PTFTEFELLNEPHGULPEKKNELLEELKATKRSIDKHTTIIOTABMGISALEKTSVP 165
    ||||| ||| ||| ||| : |||| : || | ||| : |||
Db 171 PENLFEFLYNEBPAONLTAKEKNALLPKVLKVIRESNPTRTIITDAPWMAHSAVRSKLV 230
    ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : |||
QY 186 KKEKNSIYTIHYNPFERTHOGAEWVEG---SEKWLGRKKSPPDOCKHLEEFNFIEEW 241
    || : || : || : || : || : || : || : || : || : || : || : || : || : ||
Db 231 N-DKRIIFYHYDEPKRTHOGAEWVNPPIPRVRKKNWGEENEIQLNISH---FKYSDW 285
    || : || : || : || : || : || : || : || : || : || : || : || : || : ||
QY 242 SKKKNRPIYICEFGAYRKADLESIRIKWTSFVYREMERKMSMAWYEFCSGFGYVDTLTKT 301
    : || : || : || : || : || : || : || : || : || : || : || : || : || :
Db 286 AKQNNVPPIFLEEFGAYSKADMSRVKMTESVRKMAEEFGESYAWMEFYAGFGIYDRWSQN 345
    : || : || : || : || : || : || : || : || : || : || : || : || : || :
QY 302 WNKDLLEALLIG 312
    || : || : || : || : || : || : || : || : || : || : || : || : || :
Db 346 WIEPLATAVG 356
    || : || : || : || : || : || : || : || : || : || : || : || : || :

RESULT 7
AAR49102
ID AAR49102 standard; Protein: 452 AA.
XX
XX AAR49102;
AC
XX
XX 21-JUN-1994 (first entry)
XX
XX Translated sequence of domain III of celd cDNA in clone PCNP4.
DE
XX
XX Cellulase; celd; PCNP4; anaerobic rumen.
XX
XX Neocallimastix patriciarum.
XX
XX W09400578-A.
PN
XX
XX 06-JAN-1994.
PD
XX
XX 24-JUN-1993; 93WO-A000307.
PP
XX
XX 24-JUN-1992; 92AU-0003096.
PR
XX
XX (CSIR ) COMMONWEALTH SCI & IND RES ORG.
PA
XX
XX Aylward JH, Gobius KS, Orpin CG, Xue GP.
PI
XX
XX WPI: 1994-026214/03.
DR
XX
XX N-PSDB; AA055036.
XX
XX
XX Cloning of cellulase clones from anaerobic rumen - by isolating
PT mRNA from culture, converting to cDNA etc. fungi, producing
PT enzymes useful in food processing etc., and DNA for modifying
PT rumen or silage bacteria
PT
XX
XX Claim 10; Page 46-48; 71pp; English.
PS
XX
XX Clone PCNP4.1 encodes celd. It was derived as follows. N.
CC patriciarum was used to construct a cDNA library in ZAPII (in E.
CC coli). Transformants were selected for enzyme activity using
CC selective media. 11 colonies were positive, and of these 10 had the
CC same restriction pattern, and the longest of these was designated
CC celd (PCNP4.1)(AA055036)(AAR47496,R49102). A similar process was used
CC to isolate the xylanase clone PMX-tac (AA055037,R47497). An enzyme
CC composition contg. celd and xylanase is claimed.
CC
XX
XX Sequence 452 AA;
SQ

Query Match 18.28; Score 315; DB 15; Length 452;
Best Local Similarity 27.48; Pred. No. 4.3e-20;
Matches 87; Conservative 56; Mismatches 120; Indels 54; Gaps

QY 10 KILRGINIGNALA-----PNEGDMG-VVIRDEFDIKEAGFSHVRIPR 55
    | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 18 KEMFGNMLGNTLDAQCIETYINTYKDDTASTCTCGNKNKTIEDMEFKYIMDQFNVRIPPT 77
    | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 56 WSHAVAYAPPYKINDREFFKRVDEVINGALRKGLAVAINIH-----YEELMDPEBKE 109

```

[illegible]

Query Match	18.1%	Score 314	DB 15	Length 800
Best Local Similarity	26.9%	Pred. No. 1.2e-19		
Matches 85	Conservative 52	Mismatches 127	Indels 52	Gaps 8
QY	10	KILRGINIGNALEA-----PNEGDMG-VVIRKDEFFDIKEAGFSHNRIPR	55	
DB	54	KEMNFGNMLGNTLDAQCEIYLNYDKDQTASETCWGNPKTTEDMFKVLMDNCFNRIFPTT	113	
QY	56	WSTHAYAPPPKINDREFFKRYDEVINGALKRLGAVAINHH-----YELNNDPEEHKE	109	
DB	114	WSGHFGGEAPDYKINKEMKLRVHEIYDYPKNAFVILNLHETWNHAFSETLDTAKELIE	173	
QY	110	REFLAKIOADRYKRYPETLFEELINERPGNLTPKRW-----NELLLEAKVIR	158	
DB	174	K---TWSIAIEAEFKDYDEHLIFEGLINEPRKKNTPVEMTGGDDEGMADVAMNNAVLEKTR	230	
QY	159	SI---DKKHTIIGTAEWGGSIALEKLSVPRWKNKSIVTIIHYNPFEEFTHOGAEWVEGS	214	
DB	231	SSGGNNPRRLHMLIPRYAAACNENSPKNIFFPEDDDKVIASVHNAYAPYNALNNGEQAVIDK	290	
QY	215	EKWLGRK---WCSPPDQKHLIEEFNFIEEMSKKNKRPIYIGEEGAVARKKDLERIKWTFS	271	
DB	291	FDAACKKPLEWININLMKKRFVDO-----GIPMLIGEYGAANNRDNEDBRAAMAEF	339	
QY	272	VYREMEKRRRWSNAYWE	287	
DB	340	YMEKVTAMGVPQVWMD	355	

RESULT 9
AAB08774
ID AAB08774 standard; Protein; 477 AA.

XX	02-JAN-2001	(first entry)
DT		

DE Amino acid sequence of a CE1 cellulase polypeptide.

KW
KM
KW

Cele cellulase; endoglucanase; carboxymethylcellulose; beta-glucan;
lichenin; paranitrophenyl-beta-celllobioside; cellobiose.

OS *Orpinomyces* sp.

PN US6110720-A.

PD 29-AUG-2000.

PF 17-JUL-1998; 98US-0118324.

PR 17-JUL-1998; 98US-0118324.

PA (UYGE-) UNIV GEORGIA RES FOUND INC.

PI Li X, Chen H, Ljungdahl LG;

DR WPI; 2000-586355/55.
DR N-PEDD; 2000-586355/55.

XX Neuro] and/or acti
XX

Novel nucleic acids encoding cellulase from *Orpinomyces* species PC-2
which produces cellobiose from carboxymethylcellulose, cellotetraose,
beta-glucan and lichenin -

PS Claim 2; Column 9-12; 14pp; English.

The present sequence represents a CelE cellulase. The polypeptide has endoglucanase activity, and hydrolyses carboxymethylcellulose, beta-glucan, lichenin and parrinitrophenyl-beta-D-cellobioside, to produce cellobiose. The polypeptide is derived from *Orinomyces* sp. PC-2, and shows homology to several other endoglucanases. The CelE polypeptide has a catalytic domain to glycosyl hydrolases of family 5 anaerobic bacteria. The CelE gene is devoid of introns.

[illegible]

```

RESULT 10
AAR08199
ID AAR08199 standard; protein; 406 AA.

```

AC AAR08199;

DT 01-MAR-1991 (first entry)

DE Neutral ce

KW Cellulose.

OS Ruminococcus albus.

Key	Location/Qualifiers
FM	

FT	/label= N-terminal deletion
44.400	

XX

XX
00000

XX
07-APR-1990.

XX
07-255-1000. 88 TB-0086714

XX
DA (CHTM/) CHTMTZII S

shimizu S.
XX
XX
PT

XX WPT: 1990-366379/49.
DB

DR N-PSDB; AAQ06781.
XX

PT isolated from DNA of *Ruminococcus albus* that is partly modified.

XX
PS
Claim 1: Fid 1: 11pp: Japanese.

CC The gene product is modified by a 15-24 AA N-terminal deletion and
CC a shorter C-terminal deletion to give a gene product which has

PT Piromyces rhizinflata, for producing cellulase that hydrolyzes
 PT polysaccharides containing beta-1,3 or beta-1,4 glycosidic linkages
 XX
 XX
 PS Disclosure; Columns 7-8; 16pp; English.
 CC The invention relates to a cellulase enzyme, egIA, isolated from the
 CC fungus piromyces rhizinflata. The egIA polypeptide hydrolyzes a
 CC polysaccharide containing a beta-1,3' or beta-1,4' glycosidic linkage.
 CC The egIA polynucleotide is useful for producing cellulase enzymes that
 CC can hydrolyze the glycosidic linkages in polysaccharides such as
 CC cellulose. The enzyme is also useful in a number of industrial
 CC applications where breaking down biomass is beneficial. Cellulase is also
 CC useful as a supplement in animal feed to decrease the production of fecal
 CC waste by increasing the digestibility of the feed. Cellulase is also
 CC useful to increase the efficiency of alcoholic fermentations by
 CC converting undigestible biomass into fermentable sugars. The present
 CC sequence represents an amino acid sequence of an egIA catalytic domain
 CC downstream of the GSTin an GST-egIA fusion protein.
 CC
 XX
 XX Sequence 394 AA:
 Query Match 17.8%; Score 308; DB 22; Length 394;
 Best Local Similarity 26.3%; Pred. No. 1.6e-19;
 Matches 83; Conservative 54; Mismatches 127; Indels 52; Gaps 8;
 QY 10 KILGRGINGNLEA-----PNEGDMG-VVTKDEFDIIKEAGFSHVRIPIR 55
 DB 43 KENNFQNLGNMTDACCIEYLNEKDOTASFTGCMNPKTETDMFKVLIDNQFVFRIPPT 102
 QY 56 WSHHVAFAFPPIKIMDRFEKRVDEVINGALKRGLAVAINHH-----YEELMDPEEKE 109
 DB 103 WSGHFGAPDYKIDKWLKRVHEVDYPRKNGAFVILNLHETWNAFSETLDTAKELLE 162
 QY 110 RFLAKKOIADRYKDYPTLFEFLNEPHGNLTPEKW-----NELLEBALKVIR 158
 DB 163 K--IMSQIAEEFKDYDEHLIFEGLINEPRKNDTPVEWTGDOGMDAVNANNAVFLKTVR 219
 QY 159 SI-----DKKHTITITGAEMGISALESVPKKEKNSIVTHIYNPPEFHOGAEWEGS 214
 DB 220 SAGGNPKRRLMIPYAACNENSPNNFTPEDDOKYIASVHAYPNLFALNNGGAVDK 279
 QY 215 EKMLGKR--WGSPPDQKHIEEFNFIEEMSKNRPITYIGFGAVRKADLESRIKTSF 271
 DB 280 FDAAGKRDLNEMINLMKKRRVDQ-----GIPMLIGEGYAMNRNEDRATWAEF 328
 QY 272 VVREMEKRRMSWAYVE 287
 DB 329 YMEKVTAMGVPOIWM 344
 RESULT 13
 AAB62467 standard; Protein: 582 AA.
 ID AAB62467
 XX
 AC AAB62467;
 XX
 DT 09-JUL-2001 (first entry)
 XX
 DE P. rhizinflata egIA protein partial sequence.
 XX
 XX Cellulase; egIA; fungus; polysaccharide; glycosidic linkage; enzyme;
 KW animal feed; alcohol; fermentation.
 XX
 OS Piromyces rhizinflata.
 XX
 FH Key Location/Qualifiers
 FT Region 1..109
 FT /note= "repeat region (AAB62470)"
 FT Domain 110..499
 FT /note= "catalytic domain (AAB62468)"
 FT Region 391..499
 FT /note= "repeat region (AAB62471)"
 XX

PN US6222028-B1.
 XX
 XX 24-APR-2001.
 PD
 XX
 XX 15-OCT-1999; 99US-0419459.
 PF
 XX
 XX 15-OCT-1999; 99US-0419459.
 PR
 XX
 XX (SINI-) ACAD SINICA.
 PA
 PI Liu J, Cheng K, Tsai C, Chang C;
 XX
 XX WPI: 2001-307598/32.
 DR
 XX N-PSDB: AAF83292.
 PT
 PT New isolated nucleic acid encoding a cellulase enzyme of fungus
 PT piromyces rhizinflata, for producing cellulase that hydrolyzes
 PT polysaccharides containing beta-1,3 or beta-1,4 glycosidic linkages
 CC
 PS Disclosure; Columns 5-6; 16pp; English.
 CC The invention relates to a cellulase enzyme, egIA, isolated from the
 CC fungus piromyces rhizinflata. The egIA polypeptide hydrolyzes a
 CC polysaccharide containing a beta-1,3' or beta-1,4' glycosidic linkage.
 CC The egIA polynucleotide is useful for producing cellulase enzymes that
 CC can hydrolyze the glycosidic linkages in polysaccharides such as
 CC cellulose. The enzyme is also useful in a number of industrial
 CC applications where breaking down biomass is beneficial. Cellulase is also
 CC useful as a supplement in animal feed to decrease the production of fecal
 CC waste by increasing the digestibility of the feed. Cellulase is also
 CC useful to increase the efficiency of alcoholic fermentations by
 CC converting undigestible biomass into fermentable sugars. The present
 CC sequence represents a partial egIA polypeptide.
 CC
 XX
 XX Sequence 582 AA:
 Query Match 17.8%; Score 308; DB 22; Length 582;
 Best Local Similarity 26.3%; Pred. No. 2.7e-19;
 Matches 83; Conservative 54; Mismatches 127; Indels 52; Gaps 8;
 QY 10 KILGRGINGNLEA-----PNEGDMG-VVTKDEFDIIKEAGFSHVRIPIR 55
 DB 148 KENNFQNLGNMTDACCIEYLNEKDOTASFTGCMNPKTETDMFKVLIDNQFVFRIPPT 207
 QY 56 WSHHVAFAFPPIKIMDRFEKRVDEVINGALKRGLAVAINHH-----YEELMDPEEKE 109
 DB 208 WSGHFGAPDYKIDKWLKRVHEVDYPRKNGAFVILNLHETWNAFSETLDTAKELLE 267
 QY 110 RFLAKKOIADRYKDYPTLFEFLNEPHGNLTPEKW-----NELLEBALKVIR 158
 DB 268 K--IMSQIAEEFKDYDEHLIFEGLINEPRKNDTPVEWTGDOGMDAVNANNAVFLKTVR 324
 QY 159 SI-----DKKHTITITGAEMGISALESVPKKEKNSIVTHIYNPPEFHOGAEWEGS 214
 DB 325 SAGGNPKRRLMIPYAACNENSPNNFTPEDDOKYIASVHAYPNLFALNNGGAVDK 384
 QY 215 EKMLGKR--WGSPPDQKHIEEFNFIEEMSKNRPITYIGFGAVRKADLESRIKTSF 271
 DB 385 FDAAGKRDLNEMINLMKKRRVDQ-----GIPMLIGEGYAMNRNEDRATWAEF 433
 QY 272 VVREMEKRRMSWAYVE 287
 DB 434 YMEKVTAMGVPOIWM 449
 RESULT 14
 AAM00383 standard; Protein: 574 AA.
 ID AAM00383
 XX
 AC AAM00383;
 XX
 DT 31-JAN-1997 (first entry)
 XX

DE Bacillus cellulase BCE 113.
 XX Cellulase: BCE 113; detergent; surfactant; laundry;
 KM tensile strength; antipilling.
 XX
 OS Bacillus sp. strain CBS 669.93.
 XX
 XX
 FT Key Location/Qualifiers
 FT Peptide 1..26
 FT /label= Sig-peptide
 FT 27..574
 FT Protein /label= Mat-protein
 FT /note= "The 63 kDa mature protein is used in
 detergent compns."
 XX
 XX
 XX MO9634092-AZ.
 PD 31-OCT-1996.
 XX
 XX 26-APR-1996; 96WO-EP01755.
 PF
 XX 12-MAR-1996; 96US-0614115.
 PR
 XX 28-APR-1995; 95EP-0201115.
 PR
 PA (GENV) GENENCOR INT INC.
 XX
 XX Kottwitz B, Lenting HBW, Maurer K, Van Beckhoven RFWC;
 PI Van Solingen P, Weiss A;
 XX
 DR WPI; 1996-497624/49.
 DR N-PSDB; AAT41849.
 XX
 XX Cellulase with low ratio of tensile strength loss to antipilling
 PT properties - used in detergent composition which provides
 PT anti-greying, softening, anti-wrinkling and colour protection to
 PT fabrics
 XX
 PS Claim 10; Fig 4; 33pp; English.
 XX
 XX A novel cellulase (AAW00383), designated BCE 113, of alkali-tolerant
 CC Bacillus sp. strain CBS 669.93 has a tensile strength loss to
 CC antipilling ratio below 1. It can be isolated from CBS 669.93
 CC fermentation broth or expressed at high levels in transformed host
 CC cells utilising an isolated gene sequence (AAT41849). Cellulase BCE
 CC 113, and similarly isolated cellulase BCE 103 (see also AAW00382),
 CC show good activity at alkaline pH, and can be used in laundry
 CC detergent compns to provide anti-greying, softening, anti-
 CC wrinkling and colour protection to fabrics.
 CC
 XX
 XX Sequence 574 AA;
 SQ
 Query Match 17.7%; Score 306.5; DB 17; Length 574;
 Best Local Similarity 26.0%; Pred. No. 3.6e-19;
 Matches 87; Conservative 57; Mismatches 126; Indels 65; Gaps 12;
 15 GINIGNALEA--PNEGDMG--VWIKDEFPIIKRAGESHVRIPRTWSTHAYAFPPYKIMDR 71
 45 GMLIGMTFDFAVGQDEFTAMGNPRTRELIERIADGKYSIRIPYTWENRIGADYPIDPQ 104
 72 FFRKRVDEVINGALKRGLAVAINIHHEELMNDDEERK-----ERFLALMKQIDRYKDYD 126
 105 FLNRVDEVVQWMALEBDLYVINLHDSMLWYEMEHYNGVMKYSLSMQLSNHFEDYP 164
 127 ETTFEFLNPEHGNLTPKKNELLE--EALKVIRSIDKHTIIG--TAEMGTSALEKLS 183
 165 TKLMFESVNEPKFS---QNNGETRENHHAU-----DDDLNVTVEFEYVROSQGONDTRPLV 216
 184 VPKWE-----KNSVTIHYNPFEEFTHOGAEVSESEKWLGRKWG 223
 217 LPTMERATSOPLNLNLXQITDKLDDPNLATVHYGFWPFSVNIAGYTRFE----- 268
 224 SPDDOKHLEEFNFIEBWSKKRNPYIGFCG--AYRK-----ADLESRIKWTSFVVRME 277

DB 269 - -DSKREIIEFTDRVHHTFVARGIPVVLGEGGLGPDKHTGVIOGEBKLEFFEYLHHLN 326
 OY 278 KRWSMAWMEFCSGEG-----YVDLTRKW 302
 DB 327 ERDITMLMDNGQHFNRHTYEWDEELFMDLRASW 361
 RESULT 15
 ID AAW35002
 ID AAW35002 standard; Protein; 329 AA.
 XX
 XX AAW35002;
 XX
 XX 21-MAY-1998 (first entry)
 XX
 XX Thermotoga maritima endoglucanase.
 DE
 XX Endoglucanase; cellulase; carboxymethylcellulose; cellulose;
 KM biomass; beta-1,4-glycosidic bond; hydrolysis; saccharification;
 KM thermostable enzyme; thermophilic; glycosidase.
 XX
 XX Thermotoga maritima (clone 6GA2).
 OS
 XX WO9744361-A1.
 PN
 XX 27-NOV-1997.
 PD
 XX 22-MAY-1997; 97WO-US08793.
 PF
 XX 22-MAY-1996; 96US-0651572.
 PR
 XX (RECO-) RECOMBINANT BIOCATALYSIS INC.
 PA
 XX Lam DE, Mathur EJ;
 PI WPI; 1998-018435/02.
 DR N-PSDB; AAT94210.
 XX
 XX Endoglucanase(s), preferably from archaeal bacterium, APPI 1a -
 PT useful to degrade carboxymethylcellulose and hydrolyse of
 PT beta-1,4-glycosidic bonds in cellulose
 XX
 XX Claim 1; Fig 1R; 164pp; English.
 XX
 XX This protein comprises an endoglucanase of Thermotoga maritima
 CC (clone 6GA2) that is capable of degrading carboxymethylcellulose
 CC and of hydrolysing the beta-1,4-glycosidic bonds in cellulose. It
 CC has homology to an endoglucanase of archaeobacterium AEP11a (see
 CC AAW34985). It can be produced from native cells or from recombinant
 CC host cells, especially prokaryotic host cells transformed with a
 CC plasmid or virus-derived vector including the endoglucanase DNA
 CC (see AAT94210). 24 Endoglucanases (see AAW34986-W35008) are claimed.
 CC They can be used to degrade cellulose for the conversion of plant
 CC biomass into fuels and chemicals, for use in detergents, textiles,
 CC animal feed, waste treatment, and in the fruit juice and brewing
 CC industries for the clarification and extraction of juices.
 CC
 XX
 XX Sequence 329 AA;
 SQ
 Query Match 17.2%; Score 298.5; DB 19; Length 329;
 Best Local Similarity 27.9%; Pred. No. 8.8e-19;
 Matches 94; Conservative 60; Mismatches 124; Indels 59; Gaps 16;
 14 RGINIGNALEAPNEGDMGVVIRKDEFPIIKRAGESHVRIP---IRWSTHAYAFPPYKIMD 70
 9 RGFNLLEAFSFKSTGNF---KEEDFLMAQWDFNFVRIPMCHLMDRQ---NPTIRE 61
 71 RFRKRVDEVINGALKRGLAVAINIH-----HYDELNN--DPEHKKERFLALMKQIA 119
 62 DFEKIDRVTFNCEKGIHICISLHRAFGSVNKEVEKTNLMKDETAQZAFIHHWSFIA 121
 120 DRYKDYPET-LFEELINP-----HGNLTPKKNELLEALKVIRSIDKHTIIGTAEWG 174

Db 122 RRYKGISSTHLSFNLINEPPFPDQIMSVEDHNSLIKRTITEIRKIDPERLIIIDGLGYG 181
 QY 175 GISALEKLSVPKWEKNSIVTIIHYNPFEETHOGAEWEGSE---KW---LGRKMGSP 225
 Db 182 NI-PVDDLTI---ENTVQSCRGYLPFSYTHYKAEMWDSKDPPVPEWPNMGHFGYEW--- 233
 QY 226 DDQKHLEEFNFIEWSKKNR--PIYIOFGAYRRADLESRIKWTSEVVREREMEKRRMSW 283
 Db 234 -NREKLEHY---LWIKLRQKIEVEGCGEMGAYNKTPHDVVLKWLLEDLLEIFKTLNIGF 289
 QY 284 AYWEFCGFGVYDTLRK-----TW-----NKDLLEAL 310
 Db 290 ALMNFGRPGILDSEKRDVEEEMYGHLDRKMLELL 326

Search completed: April 22, 2003, 13:57:57
 Job time : 43 secs

GenCore version 5.1.4-p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 22, 2003, 13:56:24 ; Search time 45 Seconds
(without alignments)
677.214 Million cell updates/sec

Title: US-09-880-729A-2

Sequence: 1 MGVDPEFRNKILGRGINICN.....LTKTNKDLLEALIGDSIE 317

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :
1: p1r1:*
2: p1r2:*
3: p1r3:*
4: p1r4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1728	99.8	317	B72216	endoglucanase - Th
2	598	34.5	584	J01229	cellulase (EC 3.2.
3	598	34.5	900	JH0157	cellulase (EC 3.2.
4	397	22.9	814	CZCLEM	cellulase (EC 3.2.
5	377	21.8	370	G97001	endoglucanase fam1
6	366	21.1	377	F97001	cellulase (EC 3.2.
7	362	20.9	441	A44815	cellulase (EC 3.2.
8	356.5	20.6	364	S12017	endoglucanase A
9	346.5	20.0	515	S20493	endoglucanase - Cl
10	341.5	19.7	658	A33598	endoglucanase 3 (E
11	330.5	19.1	1012	B97326	endoglucanase fam1
12	327	18.9	343	JE0409	cellulase (EC 3.2.
13	324	18.7	343	JT0268	cellulase (EC 3.2.
14	316.5	18.3	611	JC7177	endoglucanase V (E
15	315	18.2	1232	T31426	cellulase (EC 3.2.
16	311.5	18.0	409	S12018	endoglucanase B -
17	311.5	18.0	473	S40507	endoglucanase - ru
18	311	18.0	517	I40798	cellulase (EC 3.2.
19	309	17.8	406	A43722	cellulase (EC 3.2.
20	305	17.6	482	JE0302	cellulase (EC 3.2.
21	298.5	17.2	329	C72216	endo-glucanase - Th
22	297.5	17.2	574	C83725	endo-beta-1,4-gluc
23	287.5	16.6	500	S22458	cellulase (EC 3.2.
24	285.5	16.5	475	CZCLCA	cellulase (EC 3.2.
25	269.5	15.6	455	S16559	cellulase (EC 3.2.
26	265.5	15.3	566	JH0218	cellulase (EC 3.2.
27	233.5	13.5	547	J00356	cellulase (EC 3.2.
28	226.5	13.1	363	I40234	cardoxymethylcellu
29	215.5	12.4	584	S27500	ylanase - Prevote

30	206.5	11.9	397	2	A35136	cellulase (EC 3.2.
31	182	10.5	482	2	E97012	probable non-proce
32	181	10.5	409	2	B25156	cellulase (EC 3.2.
33	180	10.4	488	2	A25156	cellulase (EC 3.2.
34	173	10.0	338	2	JC4115	endo-glucanase prec
35	172.5	10.0	722	2	H96986	endo-1,4-beta gluc
36	169.5	9.8	369	2	T04323	mannan endo-1,4-be
37	165.5	9.6	915	2	A43802	cellulase (EC 3.2.
38	165.5	9.6	1039	2	S02711	cellulase (EC 3.2.
39	164.5	9.5	429	2	S29044	endoglucanase A pr
40	163.5	9.4	499	2	A27198	cellulase (EC 3.2.
41	163	9.4	403	2	A84592	(1-4)-beta-mannan
42	162.5	9.4	448	2	A27631	cellulase (EC 3.2.
43	160	9.2	410	1	S68153	cellulase (EC 3.2.
44	157.5	9.1	508	2	A26874	cellulase (EC 3.2.
45	156.5	9.0	486	2	I40548	bifunctional cellu

ALIGNMENTS

RESULT 1

B72216
endoglucanase - Thermotoga maritima (strain MSB8)

C:Species: Thermotoga maritima

C:Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 21-Jul-2000

C:Accession: B72216

R:Nelson, K.E.; Clayton, R.A.; Gill, S.R.; Gwin, M.L.; Dodson, R.J.; Haft, D.H.; Hic

Garrett, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson,

C.M.

Nature 399, 323-329, 1999

A:Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome

A:Reference number: A72200; MID:99287316; PMID:10360571

A:Accession: B72216

A>Status: Preliminary

A:Molecule type: DNA

A:Residues: 1-317 <ARN>

A:Cross-references: GB:AE001813; GB:AE000512; MID:94982321; PIDN:AAD36816.1; PID:9498

A:Experimental source: strain MSB8

C:Genetics:

A:Gene: TML751

Query Match	99.8%;	Score:1728;	DB 2;	Length 317;
Best Local Similarity	99.7%;	Pred. No. 4.5e-117;		
Matches 316;	Conservative	0;	Mismatches 1;	Indels 0;
QY	1	MGVDPEFRNKILGRGINICNALEAPNEGNGVYIKDEFDIIKEAGSHVRIPIMSTHA	60	
DB	1	MGVDPEFRNKILGRGINICNALEAPNEGNGVYIKDEFDIIKEAGSHVRIPIMSTHA	60	
QY	61	YAPPPYKIMDFRFRKRVDEVINGALRGIAVINHHVEELMNDPEEKERFLMKQIAD	120	
DB	61	YAPPPYKIMDFRFRKRVDEVINGALRGIAVINHHVEELMNDPEEKERFLMKQIAD	120	
QY	121	RYVDPEPTLFEILNEPHGNILTEPKWNLELEALKVRSIDKHTIITGAEMGISALE	180	
DB	121	RYVDPEPTLFEILNEPHGNILTEPKWNLELEALKVRSIDKHTIITGAEMGISALE	180	
QY	181	KLSVPKWEKNSIYTIHYNNFEETHOGAEVVEGSEKVLGRKWSPPDQKHLIEFPNTEE	240	
DB	181	KLSVPKWEKNSIYTIHYNNFEETHOGAEVVEGSEKVLGRKWSPPDQKHLIEFPNTEE	240	
QY	241	WSKKNRPITYIGFGAYRKADLESRIKWTSEVVREREMKRRMSNAVMEFGSGFYDRLRK	300	
DB	241	WSKKNRPITYIGFGAYRKADLESRIKWTSEVVREREMKRRMSNAVMEFGSGFYDRLRK	300	
QY	301	TWNKDLLEALIGDSIE 317		
DB	301	TWNKDLLEALIGDSIE 317		
RESULT 2				
J01229				

Query Match	34.5%;	Score 598;	DB 2;	Length 584;
Best Local Similarity	37.4%;	Pred. NO. 2e-35;		
Matches 116; Conservative	58;	Mismatches 114;	Indels 22;	Gaps 6;

RESULT 3
JH0157
cellulase (EC 3.2.1.4) H precursor - Clostridium thermocellum
N:Alternate names: endo-1,4-beta-glucanase; endoglucanase H
C:Species: Clostridium thermocellum
C:Date: 31-Dec-1991 #sequence_revision 31-Dec-1991 #text_change 15-Oct-1999
C:Accession: JH0157
R:Yagoue, E.; Beguin, P.; Aubert, J.P.

Gene 89, 61-67, 1990
A:Title: Nucleotide sequence and deletion analysis of the cellulase-encoding gene *celH* of *Trichoderma reesei*
A:Reference number: JH0157; MUID:90323606; PMID:2197182
A:Accession: JH0157
A:Molecule type: DNA
A:Residues: 1-900 <YAG>
A:Cross-references: GB:M1903; NID:g144773; PIDN:AAA3225.1; PID:g144774
A:Note: The authors translated the codon CAG for residue 863 as His
C:Comment: Cellulase H is involved in the hydrolysis of cellulose and arranged in cellulose microfibrils

Query Match	34.58;	Score 598;	DB 2;	Length 900;
Best Local Similarity	36.68;	Pred. No. 3.5e-35;		
Matches 115;	Conservative	60;	Mismatches 121;	Indels 18;
				Gaps 4

RESULT 4
CZCLEM
cellulase (EC 3.2.1.4) E precursor - Clostridium thermocellum
N:Alternate names: endo-1,4-beta-glucanase E; endoglucanase E
C:Species: Clostridium thermocellum
C:date: 31-Mar-1992 #sequence_revision 31-Mar-1992 #text_change 18-Jun-1999
C:accession: J70347
R:Hall, J.; Hazlewood, G.P.; Barker, P.J.; Gilbert, H.J.

Gene 69, 29-38 1988
A:Title: Conserved reiterated domains in Clostridium thermocellum endoglucanases are
A:Reference number: JF0347; MUID:89137992; PMID:3066698
A:Accession: JF0347
A:Molecule type: DNA
A:Residues: 1-814 <HAL>
A:Cross-references: GB:M22759; NID:q144768; PIDN:AAA23224.1; PID:q144770
A:Note: part of this sequence, including the amino end of the mature protein, was con
C:Comment: This secretory enzyme is part of a highly active and thermostable cellulase
C:Genetics:
A:Gene: cele
A:Start codon: GTG
C:Function:
A:Description: catalyzes the hydrolysis of 1,4-beta-D-glucosidic bonds in beta-D-gluc
A:Pathway: cellulose degradation
C:Superfamily: cellulase CCA; Clostridium cellulase repeat homology
C:Keywords: duplication; extracellular protein; glycosidase; hydrolase; polysaccharid
F:1-34/Domain: signal sequence #status predicted <SIC>
F:35-814/Product: cellulase E #status predicted <MAP>
F:415-438/Domain: Clostridium cellulase repeat homology <CCR1>
F:451-474/Domain: Clostridium cellulase repeat homology <CCR2>

[illegible]

RESULT 6

endoglucanase family 5 [imported] - Clostridium acetobutylicum
C:Species: Clostridium acetobutylicum
C:Date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #text_change 30-Sep-2001
C:Accession: F97001
R:Nolling, J.; Berton, G.; Omelchenko, M.V.; Matkarova, K.S.; Zeng, Q.; Gibson, R.; L.
J.; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.
J. Bacteriol. 183, 4823-4838, 2001
A>Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium
A:Reference number: A96900; MUID:21359325; PMID:21359325
A:Accession: F97001
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-577 <KUD>
A:Cross-references: GB:AE001437; PIDN:AAK78801.1; PID:g15023716; GSPDB:GN00168
A:Experimental source: Clostridium acetobutylicum ATCC6824
C:Genetics:
A:Gene: CAC0825
C:Superfamily: RumInococcus albus cellulase

Query Match 21.1%; Score 366; DB 2; Length 377;
Best Local Similarity 30.0%; Pred. No. 5,7e-19;
Matches 100; Conservative 59; Mismatches 124; Indels 50; Gaps 11;

OY 15 GINIGNALEAP-NEGDMG-----VVIKDEFDIIKEAGFSHWRIPIRMSTHAYAPPY 66
:::|||||::| | | | : | | | | | ||||| | | |
Db 55 GLDLGNSLDSPFNETDMGNPKTKSMITK-----IKQGFNTVRIPITWMSHGAPADY 108
:::|||||::| | | | : | | | | | ||||| | | |
OY 67 KIMDFEPRVDENVINGALKRGIAVINHHHEELMNDPEEKERFLA---LMQIADRY 122
:::|||||::| | | | : | | | | | ||||| | | |
Db 109 TIDNMALDRVEVVAVYADNHMYVLIIHLHEDSWLIPPYANKDATAETRLMTMQISNR 168
:::|||||::| | | | : | | | | | ||||| | | |
OY 123 KDYPEPTELFELLNEPHGMIFPEKMNELLEALAKVIRSI-----DKKHFTII 168
:::|||||::| | | | : | | | | | ||||| | | |
Db 169 KYNNSLSLFEHMNERLLGTPEEMWGCTSEARADVNSINLAANAIRDTGAKNSRFIMI 228
:::|||||::| | | | : | | | | | ||||| | | |
OY 169 GT-AEWGISALEKLVSVPKMEKNSIVTTHYYNPFEETHOGAEWESESKMLGRKGSPFD 227
:::|||||::| | | | : | | | | | ||||| | | |
Db 229 PTYAASSOKISINGLKIPNNDKILISHAYIYPNF-----AMNVNGTS-----SWGSAKD 279
:::|||||::| | | | : | | | | | ||||| | | |
OY 228 OKHLIEENFNFEEMSKKKRPRIYIGEFGARYKADLESIKRTSFVYRMREKRMSMAWE 287
:::|||||::| | | | : | | | | | ||||| | | |
Db 280 KKELDSTLNATLYKKRKGGKGVGVIGEFASVKNNMENORASLAKYVASAKRRSIAPIWMD 339
:::|||||::| | | | : | | | | | ||||| | | |
OY 288 ---FCSG---EGVYDTLRKTWN-KDLLEALIG 312
:::|||||::| | | | : | | | | | ||||| | | |
Db 340 NGVYSSGAKDSYGIFNRRNTLRMECPKIYHALTG 372
:::|||||::| | | | : | | | | | ||||| | | |

RESULT 7

A44815
cellulase (EC 3.2.1.4) EngB - Clostridium cellulovorans
N:Alternate names: endo-1,4-beta-glucanase EngB
C:Species: Clostridium cellulovorans
C:Date: 31-Mar-1993 #sequence_revision 18-Nov-1994 #text_change 09-Jun-2000
C:Accession: A44815
R:Foonng, F.; Hamamoto, T.; Shoseyov, O.; Dol, R.H.
J. Gen. Microbiol. 137, 1729-1736, 1991
A>Title: Nucleotide sequence and characteristics of endoglucanase gene engB from Clos
A:Reference number: A44815; MUID:92065240; PMID:1955860
A:Accession: A44815
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-441 <FOO>
A>Note: sequence extracted from NCBI backbone (NCBIN:68566, NCBI:P:68570)

C:Superfamily: endoglucanase; bacterial cellulose-binding domain homology
 C:Keywords: extracellular protein; glycosidase; hydrolase; polysaccharide degradation
 F:388-411/Domain: Clostridium cellulase repeat homology <CCR1>
 F:420-441/Domain: Clostridium cellulase repeat homology <CCR2>

Query Match 20.9%; Score 362; DB 2; Length 441;

Best Local Similarity 29.0%; Pred. No. 1.4e-18;
 Matches 96; Conservative 62; Mismatches 133; Indels 40; Gaps 12;

QY 15 GINIGNALEAP-NEGDMG-VVTKDEFEDLIKAGSHVRIPRMSTHVAFAFPYKIMDF 72
 C:Species: Clostridium cellulovorans
 C:Date: 19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change 16-Dec-1998
 C:Accession: S20493
 Db 52 GNLGNTMDATGGETNMGNDPLTTHAMIDVKAAGFTLPLTWDGHIAADYALDATW 111
 QY 73 FKRVEVINGALKRGALVAINIHHEELMN---DPEEKERFLALMKOIAADRYKDPET 128
 Db 112 MKRVEIATAYAFDNNMYIINLHHEGWLKPRYANAEVAKAKTKWTQIANFKYGDY 171
 QY 129 LFEELINEPHGNLTPEKW-----NELLEALKVIRSIDKHT---IITGAENG 174
 Db 172 LFEETNNEPRPVGADDEMGSGSYENRDMVNRVNLTAVNTIRATGNNALRHIVPTLAA 231
 QY 175 GIS-ALEKLSVPKWKNSIVTIHYNPFEFHGAEWEGSEKWLGRKMGSPDDOKHLE 233
 Db 232 ALSTTNDYIVPNNSRVISLHMISPYFS-----ADLTSGMTATGSDADKALSA 285
 QY 234 EFNFIEMSKKRNRPITYIGFEGAYRRADLESRIKWTSEVVRMEKRRMSWAYE--FC-- 289
 Db 286 DFDAYVNNKFVKNRAVAVIGEMGTINKNLDSEVYKAREYAKETVRGITPIMDNGCYVA 345
 QY 290 ---SGFVYDTLTK---TW-NKDLLEALIG 313
 Db 346 GKDFTEIEN--RKNLTTCCEPVMAQAFNG 374

RESULT 8

endoglucanase A - Ruminococcus albus

C:Species: Ruminococcus albus
 C:Date: 19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change 29-Sep-1999
 C:Accession: S12017
 R:Poole, D.M.; Hazlewood, G.P.; Laurie, J.I.; Barker, P.J.; Gilbert, H.J.
 Mol. Gen. Genet. 223, 217-223, 1990
 A:Title: Nucleotide sequence of the Ruminococcus albus SV3 endoglucanase genes celsA and
 A:Reference number: S12017; MUID:91066833; PMID:2250649
 A:Accession: S12017
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-364 <POO>
 A:Cross-references: GB:X54931; NID:945963; PIDN:CAA38692.1; PID:945964
 A:Note: the sequence from Fig. 4 is inconsistent with that from Fig. 2 in lacking 4-Pro.
 C:Superfamily: Ruminococcus albus cellulase

Query Match 20.6%; Score 356.5; DB 2; Length 364;

Best Local Similarity 30.3%; Pred. No. 2.7e-18;
 Matches 101; Conservative 53; Mismatches 134; Indels 45; Gaps 13;

QY 15 GINIGNALEA-----PNEGDMG-VVTKDEFEDLIKAGSHVRIPRMSTHVAFAFPYK 67
 C:Species: Fibrobacter succinogenes
 C:Date: 27-Feb-1990 #sequence_revision 27-Feb-1990 #text_change 08-Oct-1999
 C:Accession: A33598
 Db 37 GWNIGNSIDATGAPNABEVMMGNKTKEMIDAYNNKGFVIRLPVWGHVGDADPYK 96
 QY 68 IMDFRFRVDEVINGALKRGALVAINIHHEELMDPEH---KERFLALMKQIADRYK 123
 Db 97 IDDEIARVQEVVAVYADDAVVIINSHHEEDMRIPDNEHIDAVDEKTAALMKQVAREFK 156
 QY 124 DYPELLEFELINEPHGNLTPEKWNELLEALKVIRSIDK-----KHTIIG 169
 Db 157 DYGHLLIEGELNEPRVKKSPDWMNGTEGRCVRLNKTFLDTYRATGNNKELLMLT 216
 QY 170 TAEWGIS-ALEKLSVPKWKNSIVTIHYNPFEFHGAEWEGSEKWLGRKMGSPDDQ 228
 Db 217 TYASSMSNVIKDTAIPF-DHIGFSIHAYTPYAFY-----NNADWELFHMWDSHDG 269
 QY 229 KHLIEFNFIEMSKKRNRPITYIGFEGAYRRADL-ESRIKWTSEVVRMEKRR-----WS 282

Db 270 ELVSLMTNLEKENVLDKO-IPVITTEGAVNKDNNDEDRAKWSSIEYVELLGGIPCW- 327
 QY 283 WAYMEFCG---FGVYDTLTKTNKD-LLEALI 311
 Db 328 WDNGYSSGNELFGIFDRNCTWFTDTVTDALI 360

RESULT 9

endoglucanase - Clostridium cellulovorans

C:Species: Clostridium cellulovorans
 C:Date: 19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change 16-Dec-1998
 C:Accession: S20493
 R:Hamamoto, T.; Foong, F.; Shoseyov, O.; Doi, R.H.
 Mol. Gen. Genet. 231, 472-479, 1992
 A:Title: Analysis of functional domains of endoglucanases from Clostridium cellulovor
 A:Reference number: S20493; MUID:92167968; PMID:1538700
 A:Accession: S20493
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-515 <HAM>
 C:Superfamily: endoglucanase; bacterial cellulose-binding domain homology
 F:410-513/Domain: bacterial cellulose-binding domain homology <BCB>

Query Match 20.0%; Score 346.5; DB 2; Length 515;
 Best Local Similarity 27.9%; Pred. No. 2.2e-17;
 Matches 93; Conservative 62; Mismatches 137; Indels 41; Gaps 11;

QY 15 GINIGNALEA-PNEGDMG-VVTKDEFEDLIKAGSHVRIPRMSTHVAFAFPYKIMDF 72
 C:Species: Fibrobacter succinogenes
 C:Date: 27-Feb-1990 #sequence_revision 27-Feb-1990 #text_change 08-Oct-1999
 C:Accession: A33598
 Db 53 GNLGNTMDATGGETNMGNDPLTTHAMINKIKAGNTLRLVTDGGMGAAPETIADTW 112
 QY 73 FKRVEVINGALKRGALVAINIHHEELMN---DPEEKERFLALMKOIAADRYKDPET 128
 Db 113 MKRVEIATAYAFDNNMYIINLHHEGWLKPRYANAEVAKAKTKWTQIANFKYGDH 172
 QY 129 LFEELINEPHGNLTPEKW-----NELLEALKVIRSIDKHT---IITGAENG 174
 Db 173 LFEETNNEPRPVGASLEMTGSGSYENRDMVNRVNLTAVNAIRATGNNATRYIMPTLAA 232
 QY 175 GIS-ALEKLSVPKWKNSIVTIHYNPFEFHGAEWEGSEKWLGRKMGSPDDOKHLE 233
 Db 233 AMSTTNDYIVPNNSRVISLHMISPYFS-----AMDINGTS-----SMGSDYDKSSLDS 283
 QY 234 EFNFIEMSKKRNRPITYIGFEGAYRRADLESRIKWTSEVVRMEKRRMSWAYE----- 287
 Db 284 EFDAYVNNKFVKNRAVAVIGEMGTINKNNTAARVTHAEYAKSAKARGLTPIMDNGCYVA 343
 QY 288 -FCSGFVYDTLTKTN-KDLLEAL---IGDS 315
 Db 344 GKAEFTGIFNRSNLTWDAPVMAKAFITGIGSS 376

RESULT 10

endoglucanase 3 (EC 3.2.1.-) - Fibrobacter succinogenes (strain 585)

C:Species: Fibrobacter succinogenes
 C:Date: 27-Feb-1990 #sequence_revision 27-Feb-1990 #text_change 08-Oct-1999
 C:Accession: A33598
 R:McGavin, M.J.; Forsberg, C.W.; Crosby, B.; Bell, A.W.; Dignard, D.; Thomas, D.Y.
 J. Bacteriol. 171, 5587-5595, 1989
 A:Title: Structure of the cel-3 gene from Fibrobacter succinogenes 585 and characteri
 A:Reference number: A33598; MUID:90008798; PMID:2676979
 A:Accession: A33598
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-658 <MG>
 A:Cross-references: MB:M29047; NID:9148570; PIDN:AAA24893.1; PID:9148571; GB:M29681
 C:Keywords: glycosidase; hydrolase

Query Match 19.7%; Score 341.5; DB 2; Length 658;
 Best Local Similarity 26.4%; Pred. No. 6.8e-17;

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 22, 2003, 13:50:19 ; Search time 14 Seconds

(Without alignments)
939.143 Million cell updates/sec

Title: US-09-880-729A-2

Perfect score: 1732
Sequence: 1 MGVDPEERKNIKRGINIGN.....LRKTNKDLLEALIGDSIE 317

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	598	34.5	584	1 GUND_CLOCE	P25472 clostridium
2	598	34.5	900	1 GUNH_CLOTM	P16218 clostridium
3	397	22.9	814	1 GUNE_CLOTM	P10477 clostridium
4	370.5	21.4	440	1 GUNB_CLOCL	P28621 clostridium
5	356.5	20.6	364	1 GUNB_RUMAL	P23660 ruminococcus
6	341.5	19.7	658	1 GUN3_FIBSU	P14250 fibrobacter
7	338.5	19.5	515	1 GUND_CLOCL	P28623 clostridium
8	327	18.9	343	1 GUNC_CLOCE	P23340 clostridium
9	324	18.7	343	1 GUNC_CLOTM	P07985 clostridium
10	311.5	18.0	409	1 GUNB_RUMAL	P23661 ruminococcus
11	311.5	18.0	473	1 GUNB_NEOPA	P12647 neocallimastix
12	311	18.0	517	1 GUNA_CLOLO	P54937 clostridium
13	309	17.8	406	1 GUN1_RUMAL	P16216 ruminococcus
14	285.5	16.5	475	1 GUNA_CLOCE	P17901 clostridium
15	278	16.1	336	1 GUNA_RUMFL	P16169 ruminococcus
16	265.5	15.3	566	1 GUNB_PAEIA	P23550 paenibacillus
17	233.5	13.5	547	1 GUN1_BUTFI	P20847 butyrivibrio
18	201.5	11.6	397	1 GUN1_PAEPO	P23548 paenibacillus
19	181	10.5	409	1 GUN2_BACS4	P06565 bacillus
20	180	10.4	488	1 GUN1_BACS4	P06566 bacillus
21	168	9.7	400	1 GUN1_BACG4	P08545 bacillus
22	165.5	9.6	1039	1 GUNB_CALSA	P10474 c endoglucanase
23	164.5	9.5	429	1 GUNA_BUTFI	P22541 butyrivibrio
24	162.5	9.4	448	1 GUN1_CLOSA	P15704 clostridium
25	157.5	9.1	499	1 GUN1_BACS4	P07983 bacillus
26	156.5	9.0	499	1 GUN2_BACS4	P10475 bacillus
27	153	8.8	424	1 GUN1_RALSO	P10475 bacillus
28	151.5	8.7	499	1 GUN1_RALSO	P23549 ralsstonia
29	149	8.6	563	1 GUNB_CLOTM	P23549 bacillus
30	147.5	8.5	444	1 GUNB_ERMCA	P05936 erwinia
31	146.5	8.5	504	1 GUNV_ERMCA	P05935 erwinia
32	146.5	8.5	505	1 GUNV_ERMCA	P047096 erwinia
33	146	8.4	426	1 GUN2_RALSO	P17974 ralsstonia

34	144.5	8.3	466	1 GUN5_THERFU	Q01786 thermomonas
35	143.5	8.3	445	1 SPRI_YEAST	P32603 saccharomyces
36	140.5	8.1	459	1 GUNA_STRELI	P27035 streptomyces
37	139.5	8.1	418	1 GUN2_TREFE	P07982 trichoderma
38	137.5	7.9	484	1 GUNA_XANCP	P19487 xanthomonas
39	136.5	7.9	341	1 GUN1_CRYFL	Q04469 cryptococcus
40	134.5	7.8	825	1 GUN3_BACS4	P19570 bacillus
41	134	7.7	941	1 GUNB_BACS6	P19424 bacillus
42	133.5	7.7	566	1 GUNG_CLOTM	P00332 clostridium
43	129	7.4	1331	1 MANB_CALSA	P22533 caldocellum
44	124	7.2	651	1 BGLR_HUMAN	P08236 homo sapien
45	121	7.0	570	1 EXGH_SCHPO	Q10444 schizosach

ALIGNMENTS

RESULT 1
GUND_CLOCE STANDARD; PRT; 584 AA.
ID GUND_CLOCE
AC P25472;
DT 01-MAY-1992 (Rel. 22, Created)
DT 01-MAY-1992 (Rel. 22, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Endoglucanase D precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase D)
DE (cellulase D) (EGCD).
GN CELCDD.
OS Clostridium cellulolyticum.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1521;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 35319;
RX MEDLINE=92009193; PubMed=1916275;
RA Shima S., Igarashi Y., Kodama T.;
RT "Nucleotide sequence analysis of the endoglucanase-encoding gene,
RT celcdd, of Clostridium cellulolyticum.";
RL Gene 104:33-38(1991).
CC
CC -FUNCTION: THE BIOLOGICAL CONVERSION OF CELLULOSE TO GLUCOSE
CC GENERALLY REQUIRES THREE TYPES OF HYDROLYTIC ENZYMES:
CC (1) ENDOLUCANASES WHICH CUT INTERNAL BETA-1,4-GLUCOSIDIC BONDS;
CC (2) EXOCELLOBIHYDROLASES THAT CUT THE DISSACCHARIDE CELLOBIOSE
CC FROM THE NONREDUCING END OF THE CELLULOSE POLYMER CHAIN;
CC (3) BETA-1,4-GLUCOSIDASES WHICH HYDROLYZE THE CELLOBIOSE AND OTHER
CC SHORT CELLO-OLIGOSACCHARIDES TO GLUCOSE.
CC -CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic
CC linkages in cellulose.
CC -PATHWAY: Cellulose degradation.
CC -DOMAIN: A 24 RESIDUES DOMAIN IS REPEATED TWICE IN THIS ENZYME AS
CC WELL AS IN OTHER C.THERMOCELLUM CELLULOSE ENZYMES. THIS DOMAIN
CC MAY FUNCTION AS THE BINDING LIGAND FOR THE 5' COMPONENT.
CC -SIMILARITY: BELONGS TO CELLULOSE FAMILY A (FAMILY 5 OF GLYCOSYL
CC HYDROLASES).
CC
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: D90341; BAA14354.1; -
CC PIR: J01229; J01229;
CC DR HSSP: P17901; IEDG;
CC DR InterPro: IPR005087; CBM_11.
CC DR InterPro: IPR002105; Dockerin_1.
CC DR InterPro: IPR002048; EF-hand.
CC DR InterPro: IPR001547; GH_5.
CC DR Pfam: PF00150; cellulase; 1.
CC DR Pfam: PF00404; Dockerin_1; 2.
CC DR Pfam: PF03425; CBM_11; 1.

DR PROSITE: PS00018; EF_HAND; UNKNOWN_1.
 DR PROSITE: PS00448; CLOS_CELLULOSE_RPT; 1.
 DR PROSITE: PS00659; GLYCOSYL_HYDROL_F5; 1.
 DR CELLULOSE degradation; Hydrolyse; Glycosidase; Repeat; Signal.
 FT SIGNAL 1 24
 FT CHAIN 25 584
 FT DOMAIN 25 328 CATALYTIC (BY SIMILARITY).
 FT DOMAIN 329 353 PRO/THR-RICH (LINKER).
 FT DOMAIN 354 584 CELLULOSE-BINDING (BY SIMILARITY).
 FT DOMAIN 530 584 2 x 24 AA APPROXIMATE REPEATS.
 FT REPEAT 530 552 1.
 FT REPEAT 562 584 2.
 FT REPEAT 562 584 PROTON DONOR (BY SIMILARITY).
 FT ACT_SITE 159 159 NUCLEOPHILE (BY SIMILARITY).
 FT ACT_SITE 264 264 OFC41257E81322C3 CRC64;
 SO SEQUENCE 584 AA; 66061 MW; 0FC41257E81322C3 CRC64;
 Query Match 34.5%; Score 598; DB 1; Length 584;
 Best Local Similarity 37.4%; Pred. No. 4.5e-38;
 Matches 116; Conservative 58; Mismatches 114; Indels 22; Gaps 6;
 QY 10 KILGRGINIGNALPNEGDMGVVIRKDEFFDIKEAGFSHVIRIPIMSTHVAVPPYKIM 69
 DB 33 KMGIMGNIGNFEDAPTEGWSKAAQYEDDFKQAGFKHVRIPIMDQHTLANSPTVD 92
 QY 70 DFFPKVDEYINGALKRGLAVAINIHHELMNDPEHKEKRLAKOLADRYKDPET 129
 DB 93 SNEFLNIEVIMSLRGVTVYINSHDWMYSONIGREFKTIWEOIAQRKGSSENL 152
 QY 130 FFEILNEPKNITPEKNNLEALVKVIRSIDRKHITIIIGTAEWGISALEKLSVPRKWK 189
 DB 153 VEIILNEPKNITDSQINDMKNKRLINIRKTNTRVITIGAGVWNSYNSLSQLEIRP-DP 211
 QY 190 NSLVITTHVYNEPETHOGAEWEGSEKWKGRKSGPDDOKHLEEFNTEMSKKRRI 249
 DB 212 NLATFHYDYPSFTQ-----WQSGTWGTRKMDMAIMVFNHVKSDKNIPV 260
 QY 250 YIGEPAYKKADESRITKTSFVYREMERKRWMAVWEFCSG-FG-----VYDTRKT 301
 DB 261 YIGSYVGMGSHDRTSAKWFEDVSDQKISGFCGAND--NGVFGSDVNDMAYNRDTRQ 318
 QY 302 WNRDLLEAL 311
 DB 319 FDKELIINAL 328
 RESULT 2
 GUNH_CLOTM STANDARD; PRT; 900 AA.
 ID GUNH_CLOTM
 AC P16218;
 DT 01-APR-1990 (Rel. 14, Created)
 DT 01-APR-1990 (Rel. 14, Last sequence update)
 DT 15-JUL-1998 (Rel. 36, Last annotation update)
 DE Endoglucanase H precursor (EC 3.2.1.4) (EGH) (Endo-1,4-beta-glucanase) (Cellulase H)
 GN CELH.
 OS Clostridium thermocellum.
 OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
 OC Clostridium.
 OX NCBI_TaxID=1515;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC MEDLINE=90323606; Pubmed=2197182;
 RX Yagoue E., Beguin P., Aubert J.-P.;
 RA "Nucleotide sequence and deletion analysis of the cellulase-encoding
 RT gene celh of Clostridium thermocellum";
 RL Gene 89:61-67(1990).
 CC -1- FUNCTION: THIS ENZYME CATALYZES THE ENDOHYDROLYSIS OF 1,4-BETA-
 CC GLUCOSIDIC LINKAGES IN CELLULOSE, LICHENIN AND CEREAL BETA-D-
 CC GLUCANS.
 CC -1- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic
 CC linkages in cellulose.
 CC -1- DOMAIN: A 24 RESIDUES DOMAIN IS REPEATED TWICE IN THIS ENZYME AS

CC WELL AS IN OTHER C. THERMOCCELLUM CELLULOSE ENZYMES. THIS DOMAIN
 CC MAY FUNCTION AS THE BINDING LIGAND FOR THE SI COMPONENT.
 CC -1- SIMILARITY: IN THE N-TERMINAL SECTION; BELONGS TO CELLULOSE FAMILY
 CC A (FAMILY 5 OF GLYCOSYL HYDROLASES).
 CC -1- SIMILARITY: IN THE C-TERMINAL SECTION; BELONGS TO FAMILY 26 OF
 CC GLYCOSYL HYDROLASES.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: M31903; AAA23225.1; -
 DR PIR: JH0157; JH0157.
 DR HSP: P07985; ICBC.
 DR Interpro: IPR005087; CBM_11.
 DR Interpro: IPR002105; Dockerin_1.
 DR Interpro: IPR002048; EF-hand.
 DR Interpro: IPR001547; GH_5.
 DR Pfam: PF00150; cellulase_1.
 DR Pfam: PF00404; Dockerin_1; 2.
 DR Pfam: PF03425; CBM_11; 1.
 DR PROSITE: PS00018; EF_HAND; UNKNOWN_1.
 DR PROSITE: PS00448; CLOS_CELLULOSE_RPT; 2.
 DR PROSITE: PS00659; GLYCOSYL_HYDROL_F5; 1.
 DR CELLULOSE degradation; Hydrolyse; Glycosidase; Repeat; Signal.
 FT SIGNAL 1 44
 FT CHAIN 45 900
 FT DOMAIN 45 630 CATALYTIC (BY SIMILARITY).
 FT DOMAIN 631 654 PRO/THR-RICH (LINKER).
 FT DOMAIN 655 900 CELLULOSE-BINDING (BY SIMILARITY).
 FT ACT_SITE 460 460 PROTON DONOR (BY SIMILARITY).
 FT ACT_SITE 565 565 NUCLEOPHILE (BY SIMILARITY).
 FT DOMAIN 833 895 2 x 24 AA APPROXIMATE REPEATS.
 FT REPEAT 833 856 1.
 FT REPEAT 872 895 2.
 SO SEQUENCE 900 AA; 102415 MW; 973AFB1954FC246B CRC64;
 Query Match 34.5%; Score 598; DB 1; Length 900;
 Best Local Similarity 36.6%; Pred. No. 7.6e-38;
 Matches 115; Conservative 60; Mismatches 121; Indels 18; Gaps 4;
 QY 3 VDFPERKNIIGRGINIGNALPNEGDMGVVIRKDEFFDIKEAGFSHVIRIPIMSTHVA 62
 DB 327 VDFPERKMKMGKNTNIGNTLEAVYEGSWKSMKEYYEDDFKAAQYKVRIPYRMDHTR 386
 QY 63 FPPYKIMDFEKKVDEYINGALKRGLAVAINIHHELMNDPEHKEKRLAKOLADRY 122
 DB 387 TYPYTIKAFIDRVEGVVMSLSRGFTYINSHDWMKEDYNGNTEREKTIWEOIAER 446
 QY 123 KQYPEITFEIILNEPKNITPEKNNLEALVKVIRSIDRKHITIIIGTAEWGISALEK 182
 DB 447 KNSSENLFEIILNEPKNITDSQINDMKNKRLINIRKTNTRVITIGAGVWNSYNTLV 506
 QY 183 SVPRKENSIVTTHVYNEPETHOGAEWEGSEKWKGRKSGPDDOKHLEEFNTEMS 242
 DB 507 KIPD-DVYLCTFHYDYPSFTQ-----KMRG-TWGTQEDMDTVVRFEDVKSWS 554
 QY 243 KKKRPYITEFGAYKKADESRITKTSFVYREMERKRWMAVWEFCSGGVYD 296
 DB 555 DRNNIPYFGEFVMAVADRTSKVWYDFISDALEGRACSVWNGVFGSLDNDMAIYN 614
 QY 297 TLRTKNDLLEAL 310
 DB 615 RDRTFEITINAL 628
 RESULT 3
 GUNE_CLOTM STANDARD; PRT; 814 AA.
 ID GUNE_CLOTM


```

CC FT DOMAIN 387 440 2 X 24 AA APPROXIMATE REPEATS.
CC FT REPEAT 387 410 1.
CC FT REPEAT 419 440 2.
CC SQ SEQUENCE 440 AA; 48637 MW; 93DCB9EA7E2AF21 CRC64;
CC Query March 21.4%; Score 370.5; DB 1; Length 440;
CC Best Local Similarity 28.7%; Pred. No. 5,8e-21;
CC Matches 94; Conservative 62; Mismatches 137; Indels 35; Gaps 10;
CC
CC QY 15 GINIGALEAP-NEGDMG-VVIKDEFDIIKEAGFSHVRLPIRKNSTHAYAFPPRYKIMDR 72
CC | : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
CC DQ GWNLGTINTDAAGGETNMGNPLTTHAMIDKYKAAGEFNTLRPTWDGHGAAPYALADATW 111
CC 52 FKRDVEVINGLAKGLAVAINIHYYEELMN---DPDEHKERFLATKOIADRYKDYPER 128
CC QY 73 FKRDVEVINGLAKGLAVAINIHYYEELMN---DPDEHKERFLATKOIADRYKDYPER 128
CC || : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
CC DB MNRVDELNAVNYAFDNMMYIIMLHHDDCKLKRYANENAEVAKITKWTOIANPFRKVDGY 171
CC 112 MNRVDELNAVNYAFDNMMYIIMLHHDDCKLKRYANENAEVAKITKWTOIANPFRKVDGY 171
CC QY 129 LFEELLNEPHGNLPPEKM-----NELLEALVKYSIDKKHT---IIIGTVENG 174
CC | : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
CC DB 172 LIFEFMNRPGRGADDEWGSGSYENDRWVNRYNLTVAVTIRATGGNNALRHIMVPTLAAA 231
CC QY 175 GIS-ALEKLSPKMEKNSIVTIHYHNPEEPFHOGAEWEGSEKMLGRKMGSPDDOKHLIE 233
CC | : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
CC DB 232 ALSTMNDVIYPNDSRIYSLHMKSFPFS-----ADLSQTMTTWGSADNAKALSA 285
CC QY 234 EFNFIEMSKNRKPDIYGEFGAARADLESRIKWTSEFVREREMKRMSWAYWE--FC-- 269
CC : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
CC DB 286 DEDEVANKVFVKNGRAVVICEMGTIKNNLDISRVKHAERYAKEATVRCITIPMDNGCYVA 345
CC QY 290 ---SGFGVDTLRKTW-NKDLEALIGG 313
CC | : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
CC DB 346 GREQTFGIPIRNKNIITWCCEPVMAQIFRG 373
CC
CC RESULT 5
CC GUNA_RUMAL STANDARD: PRT: 364 AA.
CC ID GUNA_RUMAL
CC AC P2360;
CC DT 01-NOV-1991 (Rel. 20, Created)
CC DT 01-NOV-1991 (Rel. 20, Last sequence update)
CC DT 01-JUN-1994 (Rel. 29, Last annotation update)
CC DE Endoglucanase A (EC 3.2.1.4) (Endo-1,4-beta-glucanase) (Cellulase)
CC DE (EGC).
CC GN CELA.
CC OS Ruminococcus albus.
CC OC Bacteria; Firmicutes; Clostridia; Clostridiales; Lachnospiraceae;
CC OC Ruminococcus.
CC OX NCBI_TaxId=1264;
CC RN [1]
CC RP SEQUENCE FROM N.A., AND SEQUENCE OF 1-14.
CC RC STRAIN-SY3;
CC RX MEDLINE=91068633; PubMed=2250649;
CC RA Poole D.M., Hazlewood G.P., Laurie J.I., Barker P.J., Gilbert H.J.;
CC RT "Nucleotide sequence of the Ruminococcus albus SY3 endoglucanase
CC genes cels and cels.";
CC RL Mol. Gen. Genet. 223:217-223(1990).
CC -1- FUNCTION: HYDROLYSES BOTH CARBOXYMETHYLCELLULOSE AND XYLAN.
CC PROBABLY HAS A ROLE IN HYDROLYZING OLIGOSACCHARIDES DERIVED
CC FROM CELLULOSE, WHICH ARE TRANSPORTED ACROSS THE CELL WALL.
CC -1- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic
CC linkages in cellulose.
CC -1- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-xylosidic
CC linkages in xylans.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (Probable).
CC -1- SIMILARITY: BELONGS TO CELLULASE FAMILY A (FAMILY 5 OF GLYCOSYL
CC HYDROLASES).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sdb.ch/announce/

```

```

CC or send an email to license@slsb-sib.ch).
CC -----
DR EMBL: X54931; CAA38692.1;
DR PIR: S12017; S12017.
DR HSSP: P17901; IEDG.
DR InterPro: IPR001547; GH_5.
DR Pfam: PF00150; cellulase; 1.
DR PROSITE: PS00659; GLYCOSYL_HYDROL_F5_1.
DR Cellulose degradation: Hydrolyase; Glycosidase; Xylan degradation.
FT ACN_SITE 169 169
FT ACT_SITE 293 293
FT ACT_SITE 364 AA; 41218 MW; FBCDDE255781E7 CRC64;
SQ SEQUENCE

Query Match 20.6%; Score 356.5; DB 1; Length 364;
Best Local Similarity 30.3%; Pred. No. 5.3e-20;
Matches 101; Conservative 53; Mismatches 134; Indels 45; Gaps 13.

QY 15 GINIGNALA-----PNECDMG-VIKKEFPDIIKAGFSHVRIPIRMSTHAYAFPPYK 67
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 37 GWNIGNSLIDAGAPGNASVNMGNPKPTTKEMIDAVYNGKFDYIRLPIYMGHGDADPYK 96
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 68 IMDFPKRQVEVINGALKROLAVAINIHHEELMDPEEH-----KERPLAKQIADRYK 123
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 97 IDDEKIAVQGVVYVAVDQGAAYILINSHEEDMRIPDNEHIDAVDEKTAIWKQVAFERK 156
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 124 DYPELTFEILINERHGLTPKKNELLEALAVINSIDK-----KHTIIG 169
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 157 DYGHLLIFEGINEPRVKGSPQEMNGGTEGRRCVDRLNKTFLDIYRATGNNERILLMT 216
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 170 TAEWGGIS-ALEKLSVPRKMKNSIVTIHYNPFEEFHOGAEWEGSEKMLGRKWSPPDQ 228
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 217 TYASSMNVNKKDAIPE--DDHIGFSIHAATPYAFY-----NANADMELFHWDSDHDG 269
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 229 KHLIEEFNFIEWSKKRKPIYIGFGAATRKADL-ESRKWTSPFYRMEKRR-----WS 282
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 270 ELVSLMTMLKENYLDKID-IPVILTEGAVVKNNDNRKAKWSYIEYAEILGGLPCW- 327
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 283 MAYMEFCSG---FGYVDFLRKTKMKD-LLEALI 311
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 328 WDNGYSSGNELFGIFDRNCTWFTDTYDAII 360
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 6
GUN3_FIBSU STANDARD; PRT; 658 AA.
ID GUN3_FIBSU
AC P14250;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Endoglucanase 3 precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase)
DE (Cellulase) (Eg3).
GN CEL-3.
OS Fibrobacter succinogenes (Bacteroides succinogenes).
OC Bacteria; Fibrobacter/Acidobacteria group; Fibrobacter group;
OC Fibrobacter.
OC NCBI_TaxID=833;
OX [1]
RN SEQUENCE FROM N.A., AND SEQUENCE OF 266-287.
RP STRAIN=885.
RC MEDLINE=90008798; PubMed=2676979;
RX McGavin M.J., Forsberg C.W., Crosby B., Bell A.W., Dignard D.,
RA Thomas D.Y.;
RA "Structure of the cel-3 gene from fibrobacter succinogenes S85 and
RT characteristics of the encoded gene product, endoglucanase 3.",
RL J. Bacteriol. 171:5587-5595(1989).
CC -!- FUNCTION: EXHIBITS BOTH ENDOGLUCANASE AND CELLOBIOSIDASE
CC ACTIVITIES.
CC -!- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic
CC linkages in cellulose.
CC -!- SUBUNIT: MONOMER.
CC -!- PTM: MAY BE A LIPOPROTEIN AND MAY BE GLYCOSYLATED.
CC -!- SIMILARITY: BELONGS TO CELLULASE FAMILY A (FAMILY 5 OF GLYCOSYL
CC HYDROLASES).

```

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: M29047; AAA24893.1; -
DR PIR: A33598; A33598.
DR HSP: P07985; 1CE0.
DR InterPro: IPR005087; CBM_11.
DR InterPro: IPR001547; GH_5.
DR Pfam: PF00150; cellulase; 1.
DR Pfam: PF03425; CBM_11; 1.
DR PROSITE: PS00013; PROKAR_LIPOPROTEIN; 1.
DR PROSITE: PS00659; GLYCOSYL_HYDROL_F5; 1.
KW Cellulose degradation; Hydrolase; Glycosidase; Zymogen; Lipoprotein;
KW Signal.
FT SIGNAL 1 23 POTENTIAL.
FT PROPEP 24 265
FT CHAIN 266 658
FT LIPID 24 24
FT ACT_SITE 448 448 N-ACTY DIGLYCERIDE (POTENTIAL).
FT ACT_SITE 597 597 PROTON DONOR (BY SIMILARITY).
FT ACT_SITE 597 597 NUCLEOPHILE (BY SIMILARITY).
SQ SEQUENCE 658 AA; 73424 MW; 1C96E64C3F7109A6 CRC64;

Query Match 19.7%; Score 341.5; DB 1; Length 658;
Best Local Similarity 26.4%; Pred. No. 1.5e-18;
Matches 92; Conservative 57; Mismatches 145; Indels 55; Gaps 12;

OY 10 KIIRGINIGNALEAENEGDMGVVIRKDEFEDLIKAGFSHVRIPI---RMSTHAYAF--- 63
DB 308 KYLNKGVNTNMLENDGKFSLESDVKILADNGFKSLRPIDDLATNRDAFIAG 367
OY 64 --PPIYI-MDREFKRVDEYINGALRGALAVAINIHIEELMNDPEEKERETAL---WK 116
DB 368 TDTLKFDDDTLFLVDSFEVETAKYMSFVIDYHEYDINSYTSKADPYIKMAETWK 427
OY 117 QIADRYKQYP-ETLFEELINER---HGNTLPEKWNLEELALKVTSIDKHHIITGTA 172
DB 428 HVAHAHAEPSREDLFEELNEPDMGDKVATWTTAAQAMIDAIRTVTKHTILFGDAQ 487
OY 173 WGISALEKLSVPKMEKNSIVTTHYVNFEEFTHQGAWEV-----GSEKN----- 217
DB 488 WKSITLLAK-RIPFTDNIITVYHTEPFAFTHQGSGMTDYATIHIDIPPYDAKKSIVS 546
OY 218 --LG-----RKWGSFDDOKHLEEFNLEIEMSKKKRPYIGEGAYR-KAD 261
DB 547 GDFGVAKSTKSYVKTNIKKNYKTSKEALIEQLKAKKMAATNNVPIINIEFGALMLRST 606
OY 262 LESRIKWTSPVYVEMEKRRMSAYMEFCGFGYYDRLKTKWNDDLEAL 310
DB 607 ABRSLNLTAMREICDTLOIPWTHMGYTGNEFVIE-----NGKLIEGL 649

RESULT 7
GUND_CLOCL STANDARD; PRT; 515 AA.
AC P28623;
DT 01-DEC-1992 (rel. 24, Created)
DT 01-DEC-1992 (rel. 24, Last sequence update)
DT 01-JUN-1994 (rel. 29, Last annotation update)
DE Endoglucanase D precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase C)
DE (Cellulase D).
GN ENGD.
OS Clostridium cellulovorans.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1493;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 32-44.

RC STRAIN-ATCC 35296;
RA MEDLINE-92167968; PubMed-1538700;
RX Hamamoto T., Poong F., Shoseyov O., Dol R.H.;
RT "Analysis of functional domains of endoglucanases from Clostridium
RT cellulovorans by gene cloning, nucleotide sequencing and chimeric
RT protein construction.";
RL Mol. Gen. Genet. 231:472-479(1992).
CC -1- FUNCTION: HAS ENDOGLUCANASE ACTIVITY ON CARBOXYMETHYL-CELLULOSE
CC (CMC), CELLULOSIDASE ACTIVITY ON P-NITROPHENYL-CELLULOSE
CC (P-NPC), AND PARTIAL HYDROLYTIC ACTIVITY ON CRYSTALLINE CELLULOSE
CC (AVICEL).
CC -1- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic
CC linkages in cellulose.
CC -1- SIMILARITY: BELONGS TO CELLULOSE FAMILY A (FAMILY 5 OF GLYCOSYL
CC HYDROLASES).
CC -1- SIMILARITY: CONTAINS 1 BACTERIAL-TYPE CELLULOSE-BINDING DOMAIN
CC (CBD).

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: M37434; AAA23233.1; -
DR HSP: P07986; 1EXG.
DR InterPro: IPR001919; Bac_Cellose-bind.
DR InterPro: IPR001547; GH_5.
DR InterPro: IPR001230; Ptenyl_site.
DR Pfam: PF00150; cellulase; 1.
DR Pfam: PF00553; CBM_2; 1.
DR PROSITE: PS00659; GLYCOSYL_HYDROL_F5; 1.
KW Cellulose degradation; Hydrolase; Glycosidase; Signal.
KW Signal.
FT SIGNAL 1 31
FT CHAIN 32 515
FT DOMAIN 32 376
FT DOMAIN 377 407 CATALYTIC (BY SIMILARITY).
FT DOMAIN 408 515 PRO/THR-RICH (LINKER).
FT ACT_SITE 180 180 CELLULOSE-BINDING.
FT ACT_SITE 303 303 PROTON DONOR (BY SIMILARITY).
FT ACT_SITE 303 303 NUCLEOPHILE (BY SIMILARITY).
SQ SEQUENCE 515 AA; 55976 MW; 4CEB736CE76373F0 CRC64;

Query Match 19.5%; Score 338.5; DB 1; Length 515;
Best Local Similarity 27.6%; Pred. No. 1.9e-18;
Matches 92; Conservative 62; Mismatches 138; Indels 41; Gaps 11;

OY 15 GINGNALEA-PNEGDMG-VVIRKDEFEDLIKAGFSHVRIPIRMSTHAYAFPPYKIMDRF 72
DB 53 GWNLGMTMAIGGETMNGMNTNHAMINKIEAGFNTLRPLVWDGIMGAPEYTDQWT 112
OY 73 FKRVDEYINGALRGALAVAINIHIEELAN---DPEEKERLAIKQIADRYKQYPET 128
DB 113 MKRVEEIANAFDNDYVILNHLHENEMDKPFYANAQVAAQLTKWTOIANNEFKKGGH 172
OY 129 LFEELINEPHGNLTPEK-----NELLEALKVRSIDKHT---IITGAEWG 174
DB 173 LIFETMNERPRPVASLQWIGSGTENREYVRYLTVANAIKRAIGGNATRYINVPILAAS 232
OY 175 GIS-ALEKLSVPKMEKNSIVTTHYVNFEEFTHQGAWEVSEKWLGRKWSDDOKHLE 233
DB 233 AMSTTINDVILPNNDSKVILSHMSPYF---AMDINGTS-----SMGSDYKSLDS 283
OY 234 EFNFIEMSKKKRPYIGFGAYRKADLESRIKWTSPVYVREMEKRRMSAYME----- 287
DB 284 EFDAYVYKFKVNGRAVAVIGMGISINKNNTAARVTHAEVYAKSAKAGLPIIMWDNGYVA 343
OY 288 -FCSGGVYDTLRKTNW-KDLLEAL---IGGDS 315
DB 344 GRAETFGIFNRSLTWDAPVEMKAFIKIGGSS 376

```

RESULT 8
GUNC_CLOSF STANDARD: PRT: 343 AA.
ID GUNC_CLOSF
AC P23340:
DT 01-NOV-1991 (Rel. 20, Created)
DT 01-NOV-1991 (Rel. 20, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Endoglucanase C307 precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase)
DE (Cellulase).
GN CELC307.
OS Clostridium sp. (strain F1).
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1508;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 1-19.
RX MEDLINE=91299276; PubMed=1368690;
RA Sakka K., Shimanuki T., Shimada K.;
RT "Nucleotide sequence of celc307 encoding endoglucanase C307 of
RT Clostridium sp. strain F1."
RL Agric. Biol. Chem. 55:347-350(1991).
CC -1- FUNCTION: THIS ENZYME CATALYZES THE ENDOHYDROLYSIS OF 1,4-BETA-
CC GLUCOSIDIC LINKAGES IN CELLULOSE, LICHENIN AND CEREAL BETA-D-
CC GLUCANS.
CC -1- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic
CC linkages in cellulose.
CC -1- PATHWAY: Cellulose degradation.
CC -1- PTM: THE SIGNAL PEPTIDE WAS NOT CLEAVED IN THE PROTEIN EXPRESSED
CC IN E.COLI.
CC -1- SIMILARITY: IDENTICAL TO THE ECG FROM C.THERMOCELLUM, EXCEPT FOR
CC THREE POSITIONS.
CC -1- SIMILARITY: BELONGS TO CELLULASE FAMILY A (FAMILY 5 OF GLYCOSYL
CC HYDROLASES).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: D00945; BAA00793.1; -
CC PIR: J04009; J04009.
CC DR HSSP: P07985; 1CEN.
CC DR InterPro: IPR001347; GH_5.
CC DR Pfam: PF00150; cellulase; 1.
CC DR PROSITE: PS00659; GLYCOSYL_HYDROL_F5; 1.
CC KM Cellulose degradation; Hydrolase; Glycosidase; signal.
CC FT SIGMAL 1 21 POTENTIAL.
CC FT CHAIN 22 343 ENDOGLUCANASE C307.
CC FT ACT_SITE 140 140 PROTON DONOR (BY SIMILARITY).
CC FT ACT_SITE 280 280 NUCLEOPHILE (BY SIMILARITY).
CC SQ SEQUENCE 343 AA; 40905 MW; 72DD9BEA01A0DD05 CRC64;
CC -----
Query Match 18.9%; Score 327; DB 1; Length 343;
Best Local Similarity 28.1%; Pred. No. 8.6e-18;
Matches 97; Conservative 61; Mismatches 125; Indels 62; Gaps 13;

```

```

DB 176 YNSPELKNLA--DIDDYIYNNFHYNPEFTHQKAWSESAMAYNRYKYGQEGIE 233
QY 226 DDQKH-----LIEFNPIE-----EWSKKNRPYIIGFQAYKRADLESRI 266
DB 234 EFVKNNPKYSFMMELNNLKNLKRDLKPAIEFRERKKCKLYCGEYAIADLESRI 293
QY 267 KWTSTVREMEKRRMSMAWFECS-GFGVDTLTKRTNNKLLLEAL 310
DB 294 KWHEDIYSLBERYDIGAVWNYKKMDFEINYEDKKPVSQELVNTL 338

RESULT 9
GUNC_CLOTM STANDARD: PRT: 343 AA.
ID GUNC_CLOTM
AC P07985:
DT 01-AUG-1988 (Rel. 08, Created)
DT 01-AUG-1988 (Rel. 08, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Endoglucanase C (EC 3.2.1.4) (Egc) (Endo-1,4-beta-glucanase)
DE (Cellulase C).
GN CELC.
OS Clostridium thermocellum.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1515;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=88255856; PubMed=3384335;
RA Schwarz W.H., Schlumppig S., Ruecknagel K.P., Burschwalger S.,
RA Krell G., Staudenbauer W.L.;
RT "Nucleotide sequence of the celC gene encoding endoglucanase C of
RT Clostridium thermocellum."
RL Gene 63:23-30(1988).
RN [2]
RP ACTIVE SITE GLU-280.
RX MEDLINE=93300798; PubMed=8100226;
RA Wang Q., Tull D., Melnke A., Gillies N.R., Warren R.A., Aebertold R.,
RA Withers S.G.;
RT "Glu280 is the nucleophile in the active site of Clostridium
RT thermocellum CelC, a family A endo-beta-1,4-glucanase."
RL J. Biol. Chem. 268:14096-14102(1993).
RN [3]
RP X-RAY CRYSTALLOGRAPHY (2.15 ANGSTROMS).
RC STRAIN=NCIB 10682;
RX MEDLINE=95393242; PubMed=7664125;
RA Dominguez R., Souchon H., Spinnell S., Dauter Z., Wilson K.S.,
RA Chauvaux S., Beguin P., Alzari P.M.;
RT "A common protein fold and similar active site in two distinct
RT families of beta-glucanases."
RL Nat. Struct. Biol. 2:569-576(1995).
RN [4]
RP X-RAY CRYSTALLOGRAPHY (2.3 ANGSTROMS) OF MUTANT GLN-140.
RC STRAIN=NCIB 10682;
RX MEDLINE=96192088; PubMed=8632467;
RA Dominguez R., Souchon H., Lascombe M.-B., Alzari P.M.;
RT "The crystal structure of a family 5 endoglucanase mutant in
RT complexed and uncomplexed forms reveals an induced fit activation
RT mechanism."
RL J. Mol. Biol. 257:1042-1051(1996).
CC -1- FUNCTION: THIS ENZYME CATALYZES THE ENDOHYDROLYSIS OF 1,4-BETA-
CC GLUCOSIDIC LINKAGES IN CELLULOSE, LICHENIN AND CEREAL BETA-D-
CC GLUCANS.
CC -1- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic
CC linkages in cellulose.
CC -1- PATHWAY: Cellulose degradation.
CC -1- MISCELLANEOUS: CELC HAS AN UNUSUAL SUBSTRATE RANGE AND DISPLAYS
CC FEATURES COMMON TO CELLOBIOHYDROLASES BY BEING ABLE TO CLEAVE THE
CC ACUONIC BOND OF ARYL-BETA-GLUCOSIDES.
CC -1- SIMILARITY: IDENTICAL TO THE ECG FROM CLOSTRIDIUM SP. STRAIN F1,
CC EXCEPT FOR THREE POSITIONS.
CC -1- SIMILARITY: BELONGS TO CELLULASE FAMILY A (FAMILY 5 OF GLYCOSYL
CC HYDROLASES).
CC -----

```


CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL: M19422; AAA23220.1; -
 DR PIR: J0268; J0268.
 DR PDB: 1CEN; 03-APR-96.
 DR PDB: 1CE0; 03-APR-96.
 DR PDB: 1CEC; 29-JAN-96.
 DR InterPro: IPR001547; GH_5.
 DR Pfam: PF00150; cellulase; 1.
 DR PROSITE: PS00659; GLYCOSYL_HYDROL_F5; 1.
 DR Cellulose degradation; Hydrolase; Glycosidase; 3d-structure.
 FT ACT_SITE 140 140
 FT ACT_SITE 280 280 NUCLEOPHILE.
 FT ACT_SITE 343 343
 SQ SEQUENCE 343 AA; 40954 MW; BD24763F548C726E CRC64;

Query Match 18.7%; Score 324; DB 1; Length 343;
 Best Local Similarity 28.1%; Pred. No. 1.5e-17;
 Matches 97; Conservative 61; Mismatches 125; Indels 62; Gaps 13;

OY 15 GING---NLEAPNEGDMGVIRKDEFFDIKEAGSFHVRIPRSTHAYAFPPYKIMDR 71
 DB 7 GINGGMSISQYFSEKHFPTTEKDIETIAERGFHVALP-----FDYPIESDDN 59
 OY 72 F-----FRVDEVINGALKRGLAVAINIH-----YELMNDPEHEKRPAL 114
 DB 60 VGEYKEDGLSYDRCLCEMCKKYNGLVDMHNAQYRFQDFKSTLFEDKNQCK-REVDI 118
 OY 115 WKQIADRYKRYPELPELLEINERHGNLTPE--KNELLEALKVIRIDKKHIIITAE 172
 DB 119 WRFLAKRYINERHIAELLNEV---VEPSTRKNKLMLEYIKAIRIDSTMLYIGGN 175
 OY 173 WGISALEKLSVPEKENSIV-TIHYYNPEFTHOGAEWEG-----SEKWLGRKMGSP 225
 DB 176 YNSPDELKNLA--DIDDYIVYNHFYNPFFTHQKAMSESAMAYRYKRYQGYGIE 233
 OY 226 DDOKH-----LIEEFNIE-----EWSKKNRPIYIGEGAYRRADLESRI 266
 DB 234 EFKYNNKYSFEMLENNLKLKELLRKDLPAIEFKEKKCKLYCGEYIAIADLESRI 293
 OY 267 KMTSFYVREMEKRMWAVYEFCS-GRGVYDTLRKTNNKLLLEAL 310
 DB 294 KMHEDYISLEEDYDIGAVNNYKMKMDEIYNEDKRPVSQELVNTL 338

RESULT 10
 GUNB_RUMAL STANDARD; PRT; 409 AA.
 AC P23661;
 DT 01-NOV-1991 (Rel. 20, Created)
 DT 01-NOV-1991 (Rel. 20, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Endoglucanase B precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase)
 DE (Cellulase) (ECB).
 GN CELB.
 OS Ruminococcus albus.
 OC Bacteria; Firmicutes; Clostridia; Clostridiales; Lachnospiraceae;
 OC Ruminococcus.
 OX NCBI_TaxID=1264;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=SY3;
 RX MEDLINE=91066833; PubMed=2250649;
 RA Poole D.M., Hazlewood G.P., Laurie J.I., Barker P.J., Gilbert H.J.,
 RT "Nucleotide sequence of the Ruminococcus albus SY3 endoglucanase
 genes celsa and celB.";
 RL Mol. Gen. Genet. 223:217-223(1990).
 CC -I- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic

CC linkages in cellulose.
 CC -I- SIMILARITY: BELONGS TO CELLULASE FAMILY A (FAMILY 5 OF GLYCOSYL
 CC HYDROLASES).

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL: X54932; CAA38693.1; -
 DR PIR: S12018; S12018.
 DR HSSP: P17901; 1EDG.
 DR InterPro: IPR001547; GH_5.
 DR Pfam: PF00150; cellulase; 1.
 DR PROSITE: PS00013; PROKAR_LIPOPROTEIN; UNKNOWN_1.
 DR PROSITE: PS00659; GLYCOSYL_HYDROL_F5; 1.
 DR Cellulose degradation; Hydrolase; Glycosidase; signal.
 FT SIGNAL 1 21
 FT CHAIN 22 409
 FT ACT_SITE 212 212
 FT ACT_SITE 332 332 NUCLEOPHILE (BY SIMILARITY).
 FT ACT_SITE 409 409
 SQ SEQUENCE 409 AA; 45523 MW; 119081DFA3BFDD54 CRC64;

Query Match 18.0%; Score 311.5; DB 1; Length 409;
 Best Local Similarity 27.2%; Pred. No. 1.6e-16;
 Matches 88; Conservative 51; Mismatches 139; Indels 45; Gaps 10;

-OY 12 LGRGINGNLEAPNEG-----DW---GVYKDEFFDIKEAGSFHVRIPRSTHAYAF 63
 DB 78 MSNQMNLGNTMDATGEBLESEISWLPKYTKNFIMDLPEAGNVNLRIPVSWNHLID- 136
 OY 64 PPKYIMDFPKRDEVINGALKRGLAVAINIHVEELMNDPEEHK---ERFLAKQIAD 120
 DB 137 NNYTIDPAMMDRVOEIVNYGIDIDGMVYILNTHHEWYMPKPEKXGDIEELKAIWSDIAD 196
 OY 121 RYKDYPELPELLEINERHGNLTPEKNELLEALKVIRSIDK-----KHT 165
 DB 197 RFKGYDHLIFEGINERLRNGEGAEWGTG--TSEARELLINEKAEVETVRASGNGDRCL 255
 OY 166 IITGAEWGISALEKLSVPEKENSIVTIHYYNPEFTHOGAEWEGSEKWLGRKMGSP 225
 DB 256 MITGYAASGVNNLSALELPEDDSDKLISVHAYLPYSF-----ALDTGTGTXY-----DP 305
 OY 226 DDOKHLLIEFNFIEMSKKRPYIYIGEGAYRRADLESRIKWTSFYVREMEKRMWAV 285
 DB 306 EDTA-IPTLFESINELFISRIDPIYVGEFGSMKNIDRVCIDYDLGNNAKYDIPCVW 364
 OY 286 WEFCS-----GRGVYDTLRKTW 302
 DB 365 WDNVARIENGEBFGLNRQETDW 387

RESULT 11
 GUNB_NEOPA STANDARD; PRT; 473 AA.
 AC Q12647;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE Endoglucanase B precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase B)
 DE (Cellulase B).
 GN CELB.
 OS Neocallimastix patriciarum (Rumen fungus).
 OC Eukaryota; Fungi; Chytridiomycota; Neocallimasticales;
 OC Neocallimastixaceae; Neocallimastix.
 OX NCBI_TaxID=4758;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC MEDLINE=94128068; PubMed=8297343;
 RX Zhou L., Xue G., Orpin C.G., Black G.W., Gilbert H.J., Hazlewood G.P.;

"Ironless cell from the anaerobic fungus *Neocallimastix patriciarum* RT encodes a modular family A endoglucanase."

RL Biochem. J. 297:359-364(1994).

CC -1- FUNCTION: RATE OF HYDROLYSIS OF CELLULO-OIGOSACCHARIDES INCREASED WITH INCREASING CHAIN LENGTH FROM CELLULOSE TO CELLOPENTAOSE.

CC -1- CATALYTIC ACTIVITY: Endohydrolase of 1,4-beta-D-glucosidic linkages in cellulose.

CC -1- SIMILARITY: BELONGS TO CELLULOSE FAMILY A (FAMILY 5 OF GLYCOSYL HYDROLASES).

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC -----

DR EMBL: 231364; CAA83238.1; .

DR HSSP: P17901; IEDC.

DR InterPro: IPR002883; CBD_5.

DR InterPro: IPR001547; GH_5.

DR InterPro: IPR001230; Prenyl_site.

DR InterPro: IPR001202; WM_hsp_mmp.

DR Pfam: PF00150; cellulase; 1.

DR Pfam: PF02013; CBM_10; 2.

DR PROSITE: PS00659; GLYCOSYL_HYDROL_F5; 1.

DR PROSITE: PS01159; WM_DOMAIN_1; 1.

DR Hydrolase; Glycosidase; Repeat; Signal.

FT SIGNAL 1 17

FT CHAIN 18 473

FT DOMAIN 18 367

FT DOMAIN 367 387

FT DOMAIN 376 381

FT DOMAIN 390 473

FT REPEAT 390 427

FT REPEAT 435 473

FT ACT_SITE 173 173

FT ACT_SITE 295 295

FT ACT_SITE 473 AA; 53070 MW; B1ID9F171EA33199 CRC64;

SO SEQUENCE

Query Match 18.0%; Score 311.5; DB 1; Length 473;

Best Local Similarity 28.0%; Pred. No. 1.9e-16;

Matches 98; Conservative 53; Mismatches 136; Indels 63; Gaps 13;

QY 10 KILRGINIGALNA-----PNEGDMG-VYIKDEFDIKAGFSHVRIPIR 55

DB 30 KDLTIGSLGTLDTATCFETLDYNNKNOJASTCWNKVTQELYYKLSDLGFTFRIPPT 89

QY 56 WSTHAYAFPPYKINDREFKRVDEVINGALRGGLAVAINIH--YEELMNDPEHKEFL 112

DB 90 WSGHFGNAPDKINDQWNRVHEIYDAIINTGYALINIHETMNAHAFQKMLIESAKLLV 149

QY 113 ALMKOIAIDRYADYPTLFEFLINERHGNLTPEKW-----NELLEALKVIRSID 161

DB 150 AIWKQIAAEFDYDEHLLFEENNEPRKVGDPREMNNGGYEGENVEYENMDLFVKITRTG 209

QY 162 ----KHTITIGTAEWGSISALEKLVPRKWNKSVITVHYNPFET-HQGA---EWE 212

DB 210 GNNALRHLMITPYACINDGAINNKFPSSGDKVIVLSHVSFYFALNNGAGAISNYD 269

QY 213 GSE-KWLGKRWGSPDDQKHLIEEFNFIEMSKKKNRPYIGFGAYRKADLESRIKWTSE 271

DB 270 GSEIWM-----AMNTINSKFIISRGIPYIIGEGAMNRNEDRERMAEY 313

QY 272 VYREMEKRRWSWAYE--FCSG---FGYD--TLRKTNKDLBALIGG 313

DB 314 YIKKATISGVPCVIMDNGYFEGEGERFGLINSTIQVYVPK-LVNGLING 362

RESULT 12

GUNA_CLOLO STANDARD; PRT; 517 AA.

AC P54937;

DT 01-OCT-1996 (Rel. 34, Created)

DT 01-OCT-1996 (Rel. 34, Last sequence update)

DT 01-OCT-1996 (Rel. 34, Last annotation update)

DE Endoglucanase A precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase A) (cellulase A).

GN CELA.

OS Clostridium longisporum.

OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;

OC Clostridium.

OX NCBI_TaxID=1523;

RN [1]

RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.

RC STRAIN-ARCC 49440;

RC MEDLIN=94172316; PubMed=8126442;

RX Mittenfior V., Thomson J.A.;

RA "Cloning of an endo-(1-->4)-beta-glucanase gene, cels, from the rumen bacterium Clostridium sp. ('C. longisporum') and characterization of its product, Cels, in *Escherichia coli*."

RL J. Gen. Microbiol. 139:3233-3242(1993).

CC -1- FUNCTION: THE PH AND TEMPERATURE OPTIMA OF CELA ARE 4.8 AND 43 DEGREES CELSIUS, RESPECTIVELY. IT HYDROLYSES BARLEY BETA-GLUCAN, LICHENAN, CARBOXYMETHYLCELLULOSE AND XYLAN. IT SHOWS PREFERENTIAL ACTIVITY AGAINST THE LARGER CELLOOLIGOSACCHARIDES (CELLOHEXAOSE AND CELLOPENTAOSE); CELLOPENTAOSE IS THE SMALLEST SUBSTRATE DEGRADED COMPLETELY.

CC -1- CATALYTIC ACTIVITY: Endohydrolase of 1,4-beta-D-glucosidic linkages in cellulose.

CC -1- SIMILARITY: BELONGS TO CELLULOSE FAMILY A (FAMILY 5 OF GLYCOSYL HYDROLASES).

CC -1- SIMILARITY: CONTAINS 1 BACTERIAL-TYPE CELLULOSE-BINDING DOMAIN (CBD).

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC -----

DR EMBL: L02868; AAC37035.1; .

DR HSSP: P17901; IEDC.

DR InterPro: IPR001919; Bac_cellulose_bind.

DR InterPro: IPR001547; GH_5.

DR Pfam: PF00150; cellulase; 1.

DR PROSITE: PS00659; GLYCOSYL_HYDROL_F5; 1.

DR Cellulose degradation; Hydrolase; Glycosidase; Signal.

FT SIGNAL 1 25

FT CHAIN 26 517

FT DOMAIN 26 517

FT DOMAIN 421 517

FT ACT_SITE 185 185

FT ACT_SITE 309 309

FT ACT_SITE 517 AA; 57660 MW; A1D1570302FEBA30 CRC64;

SO SEQUENCE

Query Match 18.0%; Score 311; DB 1; Length 517;

Best Local Similarity 30.0%; Pred. No. 2.3e-16;

Matches 103; Conservative 45; Mismatches 143; Indels 52; Gaps 14;

QY 12 LGRGINIGALNA-----PNEGDMG-VYIKDEFDIKAGFSHVRIPIR 58

DB 46 MGVMGMLGNFLDAKTITMLSYNTSPISPEFTGWNPVYTKAMIDIKKAGFKTRIPPTGE 105

QY 59 HAYAFPPYKINDREFKRVDEVINGALRGGLAVAINIHVEELMNDPEEKE-----RFLA 113

DB 106 HLDG--NNKLINEWVKRKEVVDICLADDIYVILNHH--EGNVITPYAKSESVTEKLT 162

QY 114 LMKOIAIDRYADYPTLFEFLINERHGNLTPEKW-----NELLEALKVIRSI-- 160

DB 163 LMTQISEAFKDYDHLIFELINEREGTPEWTGTGTSERDVVKNYNAALESIRKGTG 222

```

OY 164 -DKKTHLIGT-AEWGGISALEKLSVPKWEKNSIYTHYNFNEFTHOGAEWEGSEKWL 218
Db 223 NNLSSAVMMPTTAAAGSGSTTMDFVPPD-DKNVIAVSAAHSYFFRA-----MDTSSNSV 275
OY 219 GRKGSPPDOKHLIEEFNFIEBMSKKNRPITIGEGFAGYRKADLESRIKWTSEVREMEK 278
Db 276 -NTWSSSDYDKXSLDVELDLYLNTFPKSGVPVYIGEGSIINKNTSSRAELAYYYTAQK 334
OY 279 RR-----WSMAYWFCSCG--FCGYPTLRKTNV-KLLEALIGC 313
Db 335 RQPCVWMDNNYAEFNKGFTGLNRSTLNMWFSIDIKALLING 377

RESULT 13
GUNI_RUMAL STANDARD: PRT: 406 AA.
AC P16216:
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 01-FEB-1995 (Rel. 31, Last annotation update)
DE Endoglucanase I precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase)
DE (Cellulase) (EG-1).
GN EG I.
OS Ruminococcus albus.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Lachnospiraceae;
OX Ruminococcus.
OX NCBI_taxid=1264;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 44-58.
RC STRAIN=F-40;
RX MEDLINE=90078126; PubMed=2687251;
RA Ohmura K., Kajino T., Kato A., Shimizu S.;
RT "Structure of a Ruminococcus albus endo-1,4-beta-glucanase gene.";
RL J. Bacteriol. 171:6771-6775(1989).
CC -I- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic
CC linkages in cellulose.
CC -I- SIMILARITY: BELONGS TO CELLULASE FAMILY A (FAMILY 5 OF GLYCOSYL
CC HYDROLASES).
CC -----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC at the European Bioinformatics Institute. There are no restrictions on
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/sib-ch).
CC or send an email to license@sib-sib.ch).
CC -----
CC CC EMBL: M30928; AAA26469.1; -.
CC PIR: A43722; A43722.
DR HSSP; P17901; 1EDG.
DR InterPro; IPR001547; GH_5.
DR Pfam; PF00150; cellulase; 1.
DR PROSITE; PS00659; GLYCOSYL_HYDROL_F5_1.
KW Cellulose degradation; Hydrolase; Glycosidase; Signal.
FT SIGNAL 1 43
FT CHAIN 44 406 ENDOGLUCANASE I.
FT ACT_SITE 210 210 PROTON DONOR (BY SIMILARITY).
FT ACT_SITE 330 330 NUCLEOPHILE (BY SIMILARITY).
FT SEQUENCE 406 AA; 45390 MW; 2E0172437B14FEAB CRC64;

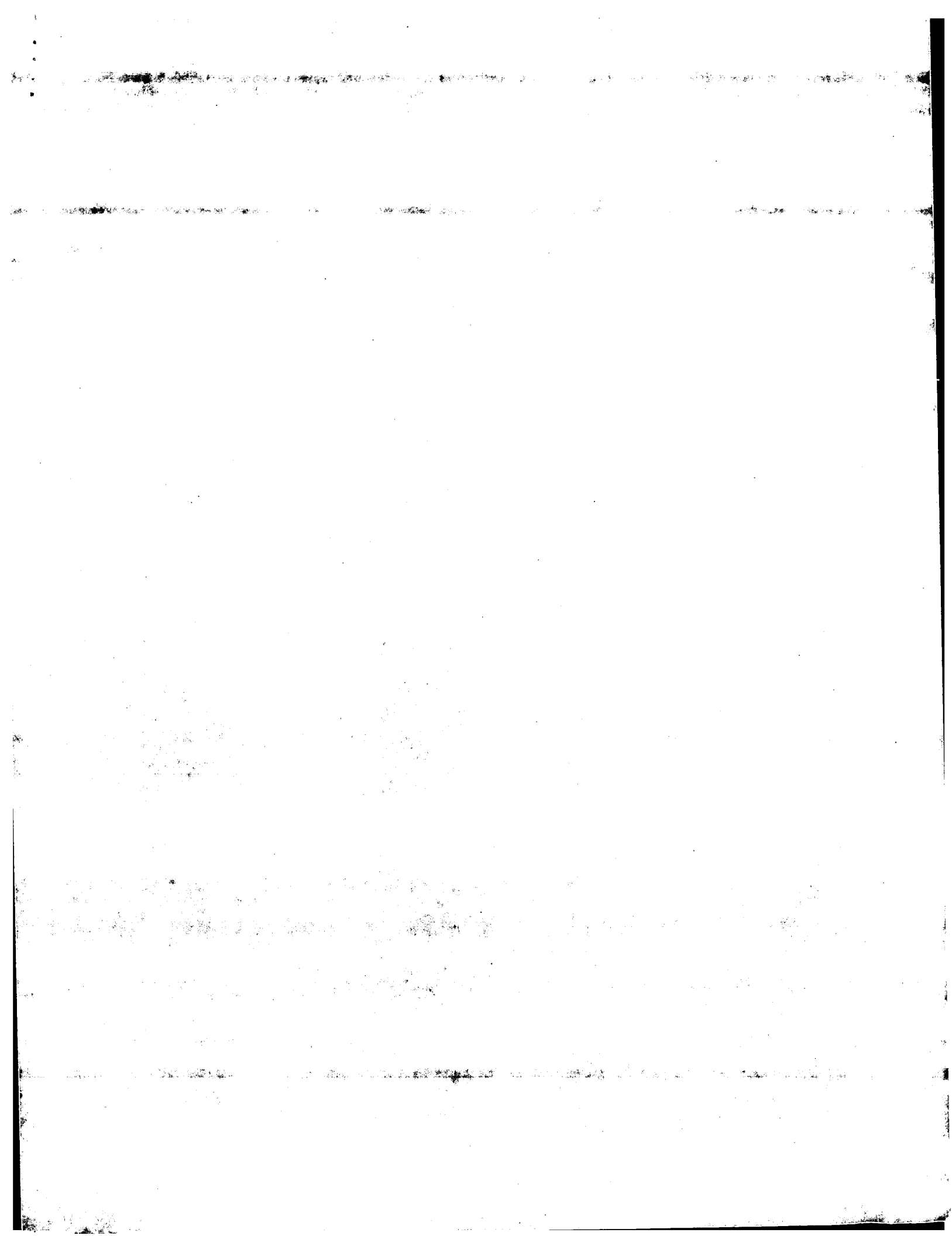
Query Match 17.8%; Score 309; DB 1; Length 406;
Best Local Similarity 27.8%; Pred. No. 2.5e-16;
Matches 92; Conservative 48; Mismatches 137; Indels 54; Gaps 12;

OY 15 GINIGNALEAFENEG-----DW---GVVAKDEFDFDIKEAGFSHVRIPIRWSFH---AYA 62
Db 79 GWNLCNTMDATAGQIGSEVSWLPKVTNNKTYIDMLPBGFVNLPIPVSWGHHIITDKYT 138
OY 63 PPKYKIMDRFRRVDEVINGALRKGLAVAINIHINHEELMNPBEERK---ERFLALMKOIA 119
Db 139 SDP-----AMMDRVOGEIYNGIDNGLVYILNTHHEEWMPKRSKMDGDEELKKAQWADIA 193
OY 120 DRYKDYPTLTFEELLINBPHGNLTPEKKMELLEALAKVIRSDK-----KH 164

```

Dd	194	DRFGYDEHLIFELINPRLRGCAEHTG-TSEARKEIINYEKAYEYVIRASGNGRC	252
Qy	165	TIIGTAEWGSISALEKLSVPKMEKNSIVTHIYNNPEFTHOGAEWEGSEKMLGRKWS	224
Dd	253	LMITGYAASSAYNNLSAIELPEDSKLLISVHAULPYSF-----ADTKGTOKY-----D	302
Qy	225	PDDCKHLEENFLFEEMSKNNKRPYIGEGGAVKAKDLDESRIKWTFSVYREKRRMSMA	284
Dd	303	PEPDA-IPLEFELHNELEFISKGPVIVGEGTMMKNEKNTEDRYKCLEDDYLAAAKYDPCV	361
Qy	285	YMEFCS-----GFGVYDLTKRMN-KDLLE	308
Dd	362	WMDNYARIGENGFGLMNRADLEWPEPDLLE	392
RESULT 14			
GN	CLDCE		
ID	GNNA_CLOCE	STANDARD;	PRT; 475 AA.
AC	PI7901;		
DT	01-NOV-1990 (Rel. 16, Created)		
DT	01-NOV-1990 (Rel. 16, Last sequence update)		
DT	01-NOV-1997 (Rel. 35, Last annotation update)		
DE	Endoglucanase A precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase A)		
DE	(Cellulase A) (EGCA).		
GN	CEUCCA.		
OS	Clostridium cellulolyticum.		
OC	Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;		
OC	Clostridium.		
OX	NCBI_TaxID=1521;		
RP	[1]		
RP	SEQUENCE FROM N.A., AND SEQUENCE OF 27-39.		
RC	STRAIN=ATCC 35319;		
RC	MEDLINE=90108715; PubMed=2558058;		
RA	Faure E., Belalich A., Bagnara C., Gaudin C., Belalich J.-P.;		
RT	"Sequence analysis of the Clostridium cellulolyticum endoglucanase-A-		
RT	encoding gene, celCCA."		
RL	[2]		
RL	gene 84:39-46(1989).		
RP	CHARACTERIZATION.		
RP	MEDLINE=92078105; PubMed=1744052;		
RA	Fierobe H.-P., Gaudin C., Belalich A., Loufti M., Faure E., Bagnara C.,		
RT	Baty D., Belalich J.-P.;		
RT	"Characterization of endoglucanase A from Clostridium		
RT	cellulolyticum."		
RL	J. Bacteriol. 173:7956-7962(1991).		
RP	[3]		
RP	MUTAGENESIS OF ARG-104; HIS-147 AND HIS-148.		
RC	MEDLINE=92325059; PubMed=1624455;		
RA	Belalich A., Fierobe H.-P., Baty D., Busetta B., Bagnara-Tardif C.,		
RA	Gaudin C., Belalich J.-P.;		
RT	"The catalytic domain of endoglucanase A from Clostridium		
RT	cellulolyticum: effects of arginine 79 and histidine 122 mutations on		
RT	catalysis."		
RL	J. Bacteriol. 174:4677-4682(1992).		
RP	[4]		
RP	X-RAY CRYSTALLOGRAPHY (1.6 ANGSTROMS) OF 27-406.		
RC	STRAIN=ATCC 35319;		
RC	MEDLINE=96097400; PubMed=8535787;		
RA	Ducros V., Czjzek M., Belalich A., Gaudin C., Fierobe H.-P.,		
RA	Belalich J.-P., Davies G.J., Haser R.;		
RT	"Crystal structure of the catalytic domain of a bacterial cellulase		
RT	belonging to family 5."		
RL	Structure 3:939-949(1995).		
CC	-1- FUNCTION: THE BIOLOGICAL CONVERSION OF CELLULOSE TO GLUCOSE		
CC	GENERALLY REQUIRES THREE TYPES OF HYDROLYTIC ENZYMES:		
CC	(1) EXOCELLULOYANASES WHICH CUT INTERNAL BETA-1,4-GLUCOSIDIC BONDS;		
CC	(2) EXOCELLULOXYDROLYASES THAT CUT THE DISSACCHARIDE CELLULOSE		
CC	FROM THE NONREDUCING END OF THE CELLULOSE POLYMER CHAIN;		
CC	(3) BETA-1,4-GLUCOSIDASES WHICH HYDROLYZE THE CELLOBIOSE AND OTHER		
CC	SHORT CELLO-OLIGOSACCHARIDES TO GLUCOSE.		
CC	-1- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic		
CC	linkages in cellulose.		

Search completed: April 22, 2003, 13:58:17
Job time : 16 secs



GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 22, 2003, 13:55:54 ; Search time 85 seconds

(without alignments)
768,435 Million cell updates/sec

Title: US-09-880-729A-2

Perfect score: 1732

Sequence: 1 MGVDPFERNRKILGRCINIGN.....LRKTNKDLLEALIGDSIE 317

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :
1: SP archaea:*
2: SP bacteria:*
3: SP fungi:*
4: SP human:*
5: SP invertebrate:*
6: SP mammal:*
7: SP mhc:*
8: SP organelle:*
9: SP phage:*
10: SP plant:*
11: SP rodent:*
12: SP virus:*
13: SP vertebrate:*
14: SP unclassified:*
15: SP viirus:*
16: SP bacteriap:*
17: SP archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1728	99.8	317	16 09X273	09X273 thermotoga
2	377	21.8	370	16 097KU0	097KU0 clostridium
3	367.5	21.2	414	2 059733	059733 rumi nococu
4	366	21.1	377	16 097KU1	097KU1 clostridium
5	363.5	21.0	637	2 059440	059440 fibrobacter
6	350.5	20.2	657	2 059446	059446 fibrobacter
7	330.5	19.1	1012	16 097DK6	097DK6 clostridium
8	328.5	19.0	332	2 060054	060054 unidentified
9	327.5	18.9	549	3 09Y870	09Y870 piromyces r
10	323	18.6	343	3 09AL36	09AL36 clostridium
11	319.5	18.4	1714	3 09P867	09P867 piromyces e
12	316.5	18.3	611	2 09LCC5	09LCC5 rumi nococu
13	315.5	18.2	669	2 059445	059445 fibrobacter
14	315	18.2	1232	3 059943	059943 neocallim
15	312	18.0	477	3 09HFE0	09HFE0 orpiniomyces
16	311.5	18.0	477	3 013333	013333 orpiniomyces

17	308	17.8	362	3 09URH5	09URH5 piromyces r
18	305	17.6	482	3 001409	001409 neocallim
19	302.5	17.5	388	3 013334	013334 orpiniomyces
20	298.5	17.2	329	16 09X274	09X274 thermotoga
21	297.5	17.2	574	16 09KF82	09KF82 bacillus ha
22	296.5	17.1	471	3 P78719	P78719 orpiniomyces
23	293	16.9	753	3 09LCC4	09LCC4 rumi nococu
24	283	16.3	400	2 008342	008342 bacillus sp
25	278	16.1	759	2 005143	005143 rumi nococu
26	277.5	16.0	356	5 09XV3	09XV3 epidi nium c
27	269.5	15.6	455	2 053302	053302 rumi nococu
28	268.5	15.5	1030	2 09XU99	09XU99 clostridium
29	264.5	15.3	322	2 09ZM63	09ZM63 prevotella
30	259.5	15.0	357	2 059441	059441 fibrobacter
31	259	15.0	519	2 047916	047916 fibrobacter
32	248.5	14.3	385	16 098F78	098F78 rhizobium l
33	226.5	13.1	353	2 044878	044878 prevotella
34	226.5	13.1	924	2 006842	006842 prevotella
35	221	12.8	1449	10 08RU51	08RU51 oryza sativ
36	215.5	12.4	584	2 045397	045397 prevotella
37	205.5	11.9	397	2 08RP23	08RP23 paenibacill
38	195	11.3	660	2 09L3J2	09L3J2 clostridium
39	190	11.0	354	2 09F0G8	09F0G8 rhizobium m
40	182	10.5	452	16 097KK6	097KK6 clostridium
41	181.5	10.5	432	2 093R81	093R81 bacillus sp
42	180	10.4	727	2 09AF65	09AF65 clavi bacter
43	179.5	10.4	424	16 098J61	098J61 rhizobium l
44	175	10.1	389	2 059232	059232 bacillus sp
45	173	10.0	368	3 012637	012637 macrophom

ALIGNMENTS

RESULT 1

ID	09X273	PRELIMINARY;	PRT;	317 AA.
AC	09X273;			
DT	01-NOV-1999 (Tremblrel. 12, Created)			
DT	01-NOV-1999 (Tremblrel. 12, Last sequence update)			
DT	01-MAR-2002 (Tremblrel. 20, Last annotation update)			
DE	ENDODUCCANASE.			
GN	TM1751.			
OS	Thermotoga maritima.			
OC	Bacteria; Thermotogae; Thermotogales; Thermotogaceae; Thermotoga.			
OX	NCBI_TaxID=2336.			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=MSB8 / DSM 3109;			
RA	MEDLINE=99287316; PubMed=10360571;			
RA	Nelson K.E., Clayton R.A., Gill S.R., Gwin M.L., Dodson R.J.,			
RA	Haft D.H., Hickey E.K., Peterson J.D., Nelson W.C., Ketchum K.A.,			
RA	McDonald L., Utterback T.R., Malek J.A., Liner K.D., Garrett M.M.,			
RA	Stewart A.M., Cotton M.D., Pratt M.S., Phillips C.A., Richardson D.,			
RA	Heidelberg J., Sutton G.G., Fleischmann R.D., Eisen J.A., White O.,			
RA	Selzberg S.L., Smith H.O., Venter J.C., Fraser C.M.;			
RT	"Evidence for lateral gene transfer between Archaea and Bacteria from			
RT	genome sequence of Thermotoga maritima.";			
RL	Nature 399:323-329(1999).			
DR	EMBL; AE001813; AAD36816.1; -			
DR	HSSP; P07985; 1CEC.			
DR	TIGR; TM1751;			
DR	InterPro; IPR001547; GH 5.			
DR	Pfam; PF00150; cellulase; 1.			
DR	PROSITE; PS00659; GLYCOSYL_HYDROL_F5; UNKNOWN_1.			
KW	Complete proteome.			
SQ	SEQUENCE 317 AA; 37383 MW; F7DF20958A9A1435 CRC64;			
QY	Query Match 99.8%; Score 1728; DB 16; Length 317;			
	Best Local Similarity 99.7%; Pred. No. 3.1e-125;			
	Matches 316; Conservative 0; Mismatches 1; Indels 0; Gaps 0;			
	1 MGVDPFERNRKILGRCINIGNALPNEGDMGVYIKDEFIDIKRAGFSHVRIPIRWSTHA 60			

```

Db      1 MGVDPEFNKILGRININGNALNEPNEGDMGVVKEDEFFDIKEAGFSVRIPIMSTHA 60
QY      61 YAPPPYKIMDRFEKRVDEVINGALKRGLAVAINIHHEELMNDPEHKEERFLAKQIAD 120
Db      61 YAPPPYKIMDRFEKRVDEVINGALKRGLAVAINIHHEELMNDPEHKEERFLAKQIAD 120
QY      121 RYKQVPELTFEFLNEPHGNLTPEKKNELLEBAKVIRSIDKRNIIIGTAMGQISALE 180
Db      121 RYKQVPELTFEFLNEPHGNLTPEKKNELLEBAKVIRSIDKRNIIIGTAMGQISALE 180
QY      181 KLSVPEKWEKNSIVTTHYNNPEFTHOGAEWVGSEKWLGRKNGSPDDOKHLEEFNTEIE 240
Db      181 KLSVPEKWEKNSIVTTHYNNPEFTHOGAEWVGSEKWLGRKNGSPDDOKHLEEFNTEIE 240
QY      241 WSKKNRPYIIGFEGAYRKADLESRIKWTSPYVREMERKRWMAWYEFSCGFVYDTLTK 300
Db      241 WSKKNRPYIIGFEGAYRKADLESRIKWTSPYVREMERKRWMAWYEFSCGFVYDTLTK 300
QY      301 TWNKDLLEALLIGDSIE 317
Db      301 TWNKDLLEALLIGDSIE 317

```

RESULT 2

```

Q97KU0      PRELIMINARY;      PRT;      370 AA.
AC      097KU0;
DT      01-OCT-2001 (TREMBlrel. 18, Created)
DT      01-OCT-2001 (TREMBlrel. 18, last sequence update)
DT      01-MAR-2002 (TREMBlrel. 20, last annotation update)
DE      Endoglucanase family 5.
GN      CAC0826.
OS      Clostridium acetobutylicum.
OC      Bacteria: Firmicutes; Bacillus/Clostridium group; Clostridia;
OC      Clostridiales; Clostridiaceae; Clostridium.
OX      NCBI_TaxID=1488;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN-ATCC 824 / DSM 792 / VKM B-1787;
RX      MEDLINE=21359325; PubMed=1146286;
RA      Noelling J., Breton G., Omelchenko M.V., Makarova K.S., Zeng Q.,
RA      Gibson R., Lee H.M., Dubois J., Qiu D., Hiltl J., Wolf Y.I.,
RA      Tatusov R.L., Sabathe F., Doucette-Stamm L., Soucaille P., Daly M.J.,
RA      Bennett G.N., Koonin E.V., Smith D.R.;
RT      "Genome sequence and comparative analysis of the solvent-producing
RT      bacterium Clostridium acetobutylicum."
RL      J. Bacteriol. 183:4823-4838(2001).
DR      EMBL: AE007598; AAK78802.1; -.
DR      InterPro: IPR001547; GH_5.
DR      Pfam: PF00150; cellulase; 1.
DR      PROSITE, PS00659; GLYCOSYL_HYDROL_F5; UNKNOWN_1.
KW      Complete proteome.
SQ      SEQUENCE 370 AA; 40934 MW; D47BDD23D2D6B92C CRC64;

```

Query Match 21.8%; Score 377; DB 16; Length 370;
 Best Local Similarity .32.2%; Pred. No. 4.2e-21;
 Matches 102; Conservative 47; Mismatches 128; Indels 40; Gaps 10;

```

QY      15 GINIGNALEA-PNEGDMG-VVKEDEFFDIKEAGFSVRIPIMSTHAYAPPPYKIMDRF 72
Db      15 GINIGNALEA-PNEGDMG-VVKEDEFFDIKEAGFSVRIPIMSTHAYAPPPYKIMDRF 72
QY      51 GWNIGNTLIDASPDDETGWGNPKTYRKAMIDRIKAGFNTRYIPVSWSHGAGSPYITIDAW 110
Db      51 GWNIGNTLIDASPDDETGWGNPKTYRKAMIDRIKAGFNTRYIPVSWSHGAGSPYITIDAW 110
QY      73 FKRQDEVINGALKRGLAVAINIHHEELMNDPEHKEERFLAKQIADRYKDYPET 128
Db      73 FKRQDEVINGALKRGLAVAINIHHEELMNDPEHKEERFLAKQIADRYKDYPET 128
QY      129 LFFELINPHGNLTPEKKNELLE-----ALKVINSIDKRNIIIGT-AMW 173
Db      129 LFFELINPHGNLTPEKKNELLE-----ALKVINSIDKRNIIIGT-AMW 173
QY      171 LIFETLNPRIAYGSPDEWNGTAESRDVINKNFLTAVNTIRGTSNNSRFLIMPTAAS 230
Db      171 LIFETLNPRIAYGSPDEWNGTAESRDVINKNFLTAVNTIRGTSNNSRFLIMPTAAS 230
QY      174 GGISLLEKLSVPEKWEKNSIVTTHYNNPEFTHOGAEWVGSEKWLGRKNGSPDDOKHLE 233
Db      174 GGISLLEKLSVPEKWEKNSIVTTHYNNPEFTHOGAEWVGSEKWLGRKNGSPDDOKHLE 233

```

```

Db      231 TATAAMDVLVPPNDRKVIYSLHWYAPYSF----AMPDKGTS-----HWGGEADKDALDG 281
QY      234 EFNFEESKKNRPYIIGFEGAYRKADLESRIKWTSPYVREMERKRWMAWYEFSCG-- 291
Db      282 QLNAYKNKFPVNGOPVYIIGEGSINKNESSRASLAKFYSDARKKGITTWMD--NGKS 339
QY      292 -----FGVYDTLTKTW 302
Db      340 AVGDNDNGILDNRNMTW 356

```

RESULT 3

```

Q59733      PRELIMINARY;      PRT;      414 AA.
AC      059733;
DT      01-NOV-1996 (TREMBlrel. 01, Created)
DT      01-NOV-1996 (TREMBlrel. 01, last sequence update)
DT      01-MAR-2002 (TREMBlrel. 20, last annotation update)
DE      Beta-1,4-D-glucanase (EC 3.2.1.4).
GN      CELA.
OS      Ruminooccus albus.
OC      Bacteria: Firmicutes; Bacillus/Clostridium group; Clostridia;
OC      Clostridiales; Lachnospiraceae; Ruminooccus.
OX      NCBI_TaxID=1264;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN-AR67;
RA      Vercore P.E., Gregg K.;
RT      "Sequence and transcriptional analysis of an endoglucanase gene from
RT      Ruminooccus albus AR67."
RL      Submitted (FEB-1993) to the EMBL/GenBank/DBJ databases.
DR      EMBL: L10243; AAA26467.1; -.
DR      HSP: P17901; IEDG.
DR      InterPro: IPR001547; GH_5.
DR      Pfam: PF00150; cellulase; 1.
DR      PROSITE, PS00659; GLYCOSYL_HYDROL_F5; UNKNOWN_1.
KW      Glycosidase; Hydrolase.
SQ      SEQUENCE 414 AA; 45879 MW; F088CFC041E3272B CRC64;

```

Query Match 21.2%; Score 367.5; DB 2; Length 414;
 Best Local Similarity 30.7%; Pred. No. 2.6e-20;
 Matches 100; Conservative 51; Mismatches 132; Indels 43; Gaps 11;

```

QY      15 GINIGNALEA-----PNEGDMG-VVKEDEFFDIKEAGFSVRIPIMSTHAYAPPPYK 67
Db      15 GINIGNALEA-----PNEGDMG-VVKEDEFFDIKEAGFSVRIPIMSTHAYAPPPYK 67
QY      68 IMDRFEKRVDEVINGALKRGLAVAINIHHEELMNDPEH-----KERFLAKQIADRYK 123
Db      68 IMDRFEKRVDEVINGALKRGLAVAINIHHEELMNDPEH-----KERFLAKQIADRYK 123
QY      146 IDENMLARQEVVYAVYDGGAVYIINSHEELRIPDNEHIDAVDEKTAIAIMKOVAERPK 205
Db      146 IDENMLARQEVVYAVYDGGAVYIINSHEELRIPDNEHIDAVDEKTAIAIMKOVAERPK 205
QY      124 DYEPETLFEFLNEPHGNLTPEKKNELLEBAKVIRSI-----DKKHTIIG 169
Db      124 DYEPETLFEFLNEPHGNLTPEKKNELLEBAKVIRSI-----DKKHTIIG 169
QY      206 DYGDHLFEFLNEPHGNLTPEKKNELLEBAKVIRSI-----DKKHTIIG 265
Db      206 DYGDHLFEFLNEPHGNLTPEKKNELLEBAKVIRSI-----DKKHTIIG 265
QY      170 T-AEWGQISALEKLSVPEKWEKNSIVTTHYNNPEFTHOGAEWVGSEKWLGRKNGSPDDQ 228
Db      170 T-AEWGQISALEKLSVPEKWEKNSIVTTHYNNPEFTHOGAEWVGSEKWLGRKNGSPDDQ 228
QY      266 TVASSCGKLQIOTPAIE-DHIGFSIHAYTYAFY-----NANADWELEFEMDNTPAT 318
Db      266 TVASSCGKLQIOTPAIE-DHIGFSIHAYTYAFY-----NANADWELEFEMDNTPAT 318
QY      229 KHLIEENFLEESKKNRPYIIGFEGAYRKADL-ESRIKWTSPYVREMERK-----WS 282
Db      229 KHLIEENFLEESKKNRPYIIGFEGAYRKADL-ESRIKWTSPYVREMERK-----WS 282
QY      319 AELITLMSNKENVYLDKIDPIVITTEYGAVCKNNNEDEDRAKWSAVLEYAEMLGIPCVW- 377
Db      319 AELITLMSNKENVYLDKIDPIVITTEYGAVCKNNNEDEDRAKWSAVLEYAEMLGIPCVW- 377
QY      283 WAYWEPGSG--FGVYDTLTKTWKND 305
Db      283 WAYWEPGSG--FGVYDTLTKTWKND 305
QY      378 WDNGYSSGNELEFJEDRNTCTWFTD 403
Db      378 WDNGYSSGNELEFJEDRNTCTWFTD 403

```

RESULT 4

```

Q97KU1      PRELIMINARY;      PRT;      377 AA.
AC      097KU1;
DT      01-OCT-2001 (TREMBlrel. 18, Created)

```


Seq	SEQUENCE	657 AA:	73554 MW:	AA29F4A0FE1B647 CAC64:
Q0	SEQUENCE	657 AA:	73554 MW:	AA29F4A0FE1B647 CAC64:
Q1	Query Match	20.2%:	Score 350.5:	DB 2: Length 657:
Q2	Best Local Similarity	27.2%:	Pred. No. 9.8e-19:	
Q3	Matches	95: Conservative	57: Mismatches	142: Indels 55: Gaps 12:
Q4	10 KILGGINCNMLEAPNEDMGVIVKDEPFDIIEKAGFSHVRIPI---	PMSTHVAFAF---	63	
Q5	308 KTLNMGVNTNTNLENAADGFKSFELGEKVOYLLAENGFSLSRPIPDLDLAIYNRDAFVWG	367		
Q6	64 --PPKYI--MDREFKRVEYINGALKRGALVAINIIHHYEELMNDPEEHKEFRLA----	WK 116		
Q7	368 TDTLTKFDDDTLFLVLDSEVWTKGNMSLVIDIHEYDMSYNTTSVKDNTYIKMAAEVK	427		
Q8	117 QIADRYKDYR-ETLPEELINEP----	HGNLTPEKMNLLLEAKLVIRSIDKKRTIITGTAE 172		
Q9	428 HVAHAHAESPRDELFEELINEPDMGKYAAQWIVAAQMDAIDITVDTKHIIITLGDQA	487		
Q10	173 MGSISALELSPVKWEKNSIVTIHYVNPFEETHOGAEWE-----	GSEKW--- 217		
Q11	488 WYSISLLAK-RPPTDQNIIVYIHTEPAFTHOGGSMEDYATIHDIPEPYDPAKSTVS	546		
Q12	218 --LG-----	RRWGSDDCKHILEENPLIEEWSKKRKPRIEFCGAYR-KAD 261		
Q13	547 GDFGVKSKRAYVKTNIKNTYKTSGEALIEQLLAKKMAATNNVPITIEFGALNLRST	606		
Q14	262 LESRIKTSFVYVEMEKRRWMAVWEFCSGGYDTLRKTNKDLLEAL	310		
Q15	607 AESRIVLTAMREICDTLQIPMTHMGYTGNSFYIE-----	NGKLIEGL 649		
Q16	RESULT 7			
Q17	ID 097DK6 PRELIMINARY: PRT: 1012 AA.			
Q18	AC 097DK6:			
Q19	DT 01-OCT-2001 (TREMBLrel. 18, Created)			
Q20	DT 01-OCT-2001 (TREMBLrel. 18, Last sequence update)			
Q21	DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)			
Q22	DE Endoglucanase family 5, S-layer homology, cell-adhesion and			
Q23	DE dockerin domains.			
Q24	CAC3469.			
Q25	OS Clostridium acetobutylicum.			
Q26	OC Bacteria: Firmicutes; Bacillus/Clostridium group; Clostridia;			
Q27	OC Clostridiales; Clostridiaceae; Clostridium.			
Q28	OX NCBI_TaxID=1488:			
Q29	11			
Q30	RP SEQUENCE FROM N.A.			
Q31	RC STRAIN=ATCC 824 / DSM 792 / VKM B-1187;			
Q32	RX MEDLINE=11359325; Pubmed=11466286;			
Q33	RA Noelling J., Breton G., Omelchenko M.V., Makarova K.S., Zeng Q.,			
Q34	RA Gibson R., Lee H.M., Dubois J., Qiu D., Hitti J., Wolf Y.I.,			
Q35	RA Tatusov R.L., Sebaste F., Doucetste-stam L., Soucaille P., Daly M.J.,			
Q36	RA Bennett G.N., Koonin E.V., Smith D.R.,			
Q37	RT "Genome sequence and comparative analysis of the solvent-producing			
Q38	RT bacterium Clostridium acetobutylicum.";			
Q39	RL J. Bacteriol. 183:4823-4838(2001).			
Q40	EMBL: AE007844: AAK81397.1: -			
Q41	DR InterPro: IPR003343: Btg_2.			
Q42	DR InterPro: IPR002105: Dockerin_1.			
Q43	DR InterPro: IPR002048: EF-hand.			
Q44	DR InterPro: IPR001547: GH_5.			
Q45	DR InterPro: IPR001950: TIF_SUI1.			
Q46	DR Pfam: PF02368: Btg_2; 1.			
Q47	DR Pfam: PF00150: cellulase; 1.			
Q48	DR Pfam: PF00404: Dockerin_1; 2.			
Q49	DR PROSITE: PS00448: CLOS_CELULOSOME_RPT; UNKNOWN_1.			
Q50	DR PROSITE: PS00018: EF_HAND; UNKNOWN_2.			
Q51	DR PROSITE: PS00659: GLYCOSYL_HYDROL_F5; UNKNOWN_1.			
Q52	DR PROSITE: PS01118: SUI1_1; UNKNOWN_1.			
Q53	KM Complete proteome.			
Q54	Q0 SEQUENCE 1012 AA: 110097 MW: A74C99E6CFCE945 CRC64:			

```

Best Local Similarity 28.3%; Pred No. 6e-17;
Matches 98; Conservative 66; Mismatches 131; Indels 51; Gaps 16;

QY      2 GVDPERKKILIGRINIGNALE-APNEGDMG-VVIKDEFDILKEAGFSHVRPIRKSTH 59
      1 : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      383 GLSSNDILVKDKMGVMDGNLSLESAGGEGMGNPATRTKMKIDOMKKAQFNVRIPRWDEH 442
      1 : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      60 AYAPRPYKIMRFRKRVDEVINGALRKGLAIVAHINHYEELMDNPEEKREFL-ALW 115
      1 : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      443 -YTSNATIDPAIRYSRVTVVNYVALANGMAVATIHINHKLOGEAMNAKMSVINEGSAIW 501
      1 : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      116 KQIADRYKDYETELFEELLNEPHGNTLPEKW-----NELLEALKYINST---DK 162
      1 : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      502 KQVGHGFEDYDQGLFETINEPR---TGDDMTGNSSYYNVNNEFNAKTLSTIRATGNGG 558
      1 : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      163 KHTITIGF----AEMGGISALEKLSVPRWEKNSIVTIHYNPFETTHOGAEWVGSEKWL 218
      1 : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      559 KRLVMPYPCASADAKAAV--MVNP-DRVAVSHATIPYVFA-MNDPTTGSGVSTF 613
      1 : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      219 GRKMGSPDDOKHLIEE-FNFIEMSKNRKRPYIGEGAVRKADLESRIKWTSPVREME 277
      1 : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      614 G-----DADKACIDKTELLDKTEFKIGIPVVGEEFATCRKNLODVPNAKYLLQVAS 667
      1 : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      278 KRKMSAWNE-----FCSGRGVDTLRKTMN--KDLLEALIGG 313
      1 : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      668 SYGIPCCWMDNNAESPNTSDSMGLFN--RKTLQFYFPELVQAMLDG 711
      1 : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 8
060054 PRELIMINARY: PRT: 332 AA.
ID 060054
AC 060054
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)
DE Cellulase (EC 3.2.1.4).
GN CEL.
OS unidentified bacterium.
OC Bacteria; environmental samples.
OX NCBI_TaxID=2338;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96045463; PubMed=7546604;
RA Healy F.G., Ray R.M., Aldrich H.C., Wolkie A.C., Ingram L.O.,
RA Shannugam K.T.;
RT "Direct isolation of functional genes encoding cellulases from the
RT microbial consortia in a thermophilic, anaerobic digester maintained
RT on lignocellulose.";
RL Appl. Microbiol. Biotechnol. 43:667-674(1995).
DR EMBL: U12011; AAA91966.1; -.
DR HSSP: P07985; ICEN.
DR InterPro: IPR001547; GH_5.
DR Pfam: PF00150; cellulase; 1.
KW Glycosidase; Hydrolase.
SQ
SEQUENCE 332 AA; 38429 MW; 3D34FA90C39129C1 CRC64;

*Query Match 19.0%; Score 328.5; DB 2; Length 332;
Best Local Similarity 27.1%; Pred. No. 2e-17;
Matches 92; Conservative 52; Mismatches 135; Indels 61; Gaps 11.

Db      9 RGINIGMISQCPREOSHDSFTIGEDIKITASMGDHIKLP-----DYDVSD 58
      1 : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      14 RGINIGNAL-EAANE-GDMGVATKDEFDITKAGFSHVRPIRWSTHAYAPRPYK- 68
      1 : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      9 RGINIGMISQCPREOSHDSFTIGEDIKITASMGDHIKLP-----DYDVSD 58
      1 : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      69 -----MDFFRRVDEVINGALRKGLAIVAHINTH-----HYEELMDNPEEKREFLALMK 116
      1 : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      59 DEGLTIGIKIVYDNCVWCNRNGNLTLIDHKAPGYAFYVNEEDWTLTFESRVDQAFVNLK 118
      1 : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      117 QIADRYKDYETELFEELLNEPHGNTLPEKKNELLEFLKVIKRSIDKKHTIIGNAEWGI 176
      1 : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      119 EFARVRYKYSDFVSEFELLENEV-VDADGLMNDLIRRTITELRTAPRTKAYVVGIRKNSA 177
      1 : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      177 SALEKLSVPRWEKNSIVTIHYNPFETTHOGAEWVGSEKWLGRKMGSPDDOKHLIEEFN 236
      1 : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```


Db 483 KKHILFGDAQWYISILIAK-RTPFTDDNIYVITYHPFAFTHQGSMTDYATIKGIPF 541
QY 213 --GSEKW-----LG-----RKWSPDDOKHLEEFNFIEEMSKKRPYI 251
Db 542 PYDAKSTYSGDGVKSTQSYKSNIKNYKSGKEALIVEILKAKKMAATNNVPII 601
QY 252 GEFCAVR-KADLESRIKTSFVREMEKRRMSWAYWECGSGFYVD 296
Db 602 NEFALNLRSTADRLNLYFTAMREICDTLQIPWTHWGYTGWFSLE 647

RESULT 14

059943 PRELIMINARY; PRT; 1232 AA.
AC 059943;
DT 01-AUG-1998 (Tremblrel. 07, Created)
DT 01-AUG-1998 (Tremblrel. 07, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE Cellulase Celld (Fragment).
GN CEID.
OS Neocallimastix patriciarum (Rumen fungus).
OC Eukaryota; Fungi; Chytridiomycota; Neocallimasticales;
OC Neocallimasticeae; Neocallimastix.
OX NCBI_TaxID=4758;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93123992; Pubmed=1479358;
RUE G.P., Gobius K.S., Orpin C.G.;
RT "A novel polysaccharide hydrolase cDNA (celld) from Neocallimastix patriciarum encoding three multi-functional catalytic domains with high endoglucanase, cellobiohydrolase and xylanase activities.";
RL J. Gen. Microbiol. 138:0-0(0).
DR EMBL; AF053363; AAC06321.1; -;
DR HSSP; P17901; 1EDG.
DR InterPro; IPR002883; CBD_5.
DR InterPro; IPR001547; GH_5.
DR Pfam; PF02013; CBM_10; 2.
DR Pfam; PF00150; cellulase; 3.
DR PROSITE; PS00659; GLYCOSYL_HYDROL_F5; UNKNOWN_3.
FT NON TER 1
SQ SEQUENCE 1232 AA; 140617 MW; E58133110F684348 CRC64;

Query Match 18.2%; Score 315; DB 3; Length 1232;
Best Local Similarity 27.4%; Pred. No. 1.2e-15;
Matches 87; Conservative 56; Mismatches 120; Indels 54; Gaps 10;

QY 10 KILRGINIGNALEA-----PNEGDMG-VINDEFDIIKEAGFSHVRIPIR 55
Db 798 KEMFEGNNLGNLTDAOCIEYLNIDKDOTASETCGNPKTTEDEKVLMDNQFNFRIPPT 857
QY 56 WSTHAVALPPYKIMDRFFKRVDEVINGALRKGLAVAINIH-----YEELMNDPEEHE 109
Db 858 WSGHFGAPRYKINEKLVKHEIVDYPKNGAFVILNLHETMNAFSETLDIAKILE 917
QY 110 RFLALMQLADRYKYDETELEFELINPHGNLPPEKVN-----NELLEALKVIR 158
Db 918 K---IWSQIAEEFDYDEHILFEGLINPRKNDPEVETGDDQEGDAVNMNNAVFLKTR 974
QY 159 SI-----DKKTIIGTLEMWGISALEKLSVPMKMNKSVIIVHYNPEFT-HOGAEVVEG 213
Db 975 SSGGNPKKRLMLPYTAACNENSKNFPEDDDKVIASVHAAPYALNNGAGAVND 1034
QY 214 SEKWLGKR--WCSPPDOKHLEEFNFIEEMSKKRPYIGFCAVYRKADLESRIKTS 270
Db 1035 FDA-AGKKDLEWMINILMKKRFVDQ-----GIFMLIGEGAMNRRNEERATVAE 1082
QY 271 FVYREMEKRRMSWAYWE 287
Db 1083 FYMEKVTAMGVPOVWMD 1099

RESULT 15
Q9HFE0

ID Q9HFE0 PRELIMINARY; PRT; 477 AA.
AC Q9HFE0;
DT 01-MAR-2001 (Tremblrel. 16, Created)
DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE 1,4-Delta-D-glucan-4-glucanohydrolase (EC 3.2.1.4).
GN CELE.
OS Orphiomyces sp. PC-2.
OC Eukaryota; Fungi; Chytridiomycota; Neocallimasticales;
OC Neocallimasticeae; Orphiomyces.
OX NCBI_TaxID=50059;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PC-2;
RX MEDLINE=98146543; Pubmed=9485595;
RA Chen H., Li X.L., Blum D.L., Ljungdahl L.G.;
RT "Two genes of the anaerobic fungus Orphiomyces sp. strain PC-2 encoding cellulases with endoglucanase activities may have arisen by gene duplication.";
RL FEMS Microbiol. Lett. 159:63-68(1998).
DR EMBL; U97153; AAC05164.1; -;
DR HSSP; P17901; 1EDG.
DR InterPro; IPR002883; CBD_5.
DR InterPro; IPR001547; GH_5.
DR InterPro; IPR001230; Premyl_site.
DR Pfam; PF02013; CBM_10; 2.
DR Pfam; PF00150; cellulase; 1.
DR PROSITE; PS00659; GLYCOSYL_HYDROL_F5; UNKNOWN_1.
DR PROSITE; PS00294; PRENITATION; UNKNOWN_1.
KW Glycosidase; Hydrolase.
SQ SEQUENCE 477 AA; 53635 MW; 6CBEC85A469BA0DA CRC64;

Query Match 18.0%; Score 312; DB 3; Length 477;
Best Local Similarity 27.9%; Pred. No. 5.9e-16;

Matches 96; Conservative 49; Mismatches 115; Indels 84; Gaps 14;

QY 10 KILRGINIGNALRAP--NEGDMGVIRKDE-----FPDIKEAGFSHVRI 52
Db 31 KELLIGNNGLALDAHGLDKLDYV--KDQLASETCNANPKATPGLSALKNGQFNFR 87
QY 53 PIRKSTHAVALPPYKIMDRFFKRVDEVINGALRKGLAVAINIHYE--ELMNDPEEHE 109
Db 88 PTTWTGHRGNGPDYKIDVMMRRVHEVVDYALNTGSVYIINIHENNNYAFSNNLOKAKP 147
QY 110 RFLALMQLADRYKYDETELEFELINPHGNLPPEKVNLEELAKVIRSDKKHTIIG 169
Db 148 ILAALIMQLAERFANYDEHILFEGLINPRKNDPEVETGDDQEGDAVNMNNAVFLKTR 204
QY 170 TAE-WGISALEKLSVPMKMNKSVIIVHYNPEFT--THQ 206
Db 205 TVRASGNNMIRHLMPTIYACVNDGALSIVYKFTPTNDKVIASVSYVPYNFALNTGA 264
QY 207 GAENVESSEKWLGRKWSPPDOKHLEEFNFIEEMSKK-KR-----PIYIGFCAVYR 258
Db 265 GAE-----KTFGSTSD-----IEMANNIKRFLVDRNIPVIGIFEGAMN 303
QY 259 KADLESRIKTSFVREMEKRRMSWAYWE--FCSG-----GVVD 296
Db 304 RDNESERARNAEYIKSATATAMGVPCVILMDNGYTOGTGELFGLVD 347

Search completed: April 22, 2003, 13:59:51
Job time : 89 secs

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 22, 2003, 13:57:14 ; Search time 29 Seconds

(without alignments)
321.623 Million cell updates/sec

Title: US-09-880-729A-2

Sequence: 1 MGVDPERNKILRGINIGIN.....LRTKWNKDLLEALIGDSIE 317

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

Issued_Patents_AA:*
1: /cgn2_6/prodata/1/1aa/5A_COMB.pep:*
2: /cgn2_6/prodata/1/1aa/5B_COMB.pep:*
3: /cgn2_6/prodata/1/1aa/6A_COMB.pep:*
4: /cgn2_6/prodata/1/1aa/6B_COMB.pep:*
5: /cgn2_6/prodata/1/1aa/6COTUS_COMB.pep:*
6: /cgn2_6/prodata/1/1aa/backfillsl.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1732	100.0	317	2	US-09-066-075-2
2	1732	100.0	317	2	US-08-518-615A-2
3	1732	100.0	317	2	US-08-951-889-2
4	1732	100.0	317	4	US-09-472-857-2
5	835.5	48.2	360	4	US-09-134-078-27
6	312	18.0	477	3	US-09-118-324-2
7	311.5	18.0	473	4	US-09-286-691-15
8	311.5	18.0	473	4	US-09-687-147-15
9	308	17.8	390	4	US-09-419-459-4
10	308	17.8	394	4	US-09-419-459-10
11	308	17.8	582	4	US-09-419-459-2
12	306.5	17.7	574	4	US-08-732-433-1
13	306.5	17.7	574	4	US-08-945-574-2
14	296.5	17.1	471	3	US-09-118-324-5
15	296.5	17.1	471	4	US-09-286-691-12
16	296.5	17.1	471	4	US-09-687-147-12
17	281.5	16.3	537	4	US-09-110-959A-4
18	265.5	15.3	466	2	US-07-862-588B-4
19	177	10.2	467	2	US-08-727-548-2
20	177	10.2	467	4	US-08-945-574-1
21	170	9.8	352	4	US-09-102-204-1
22	168	9.7	400	2	US-08-713-298B-2
23	168	9.7	400	3	US-08-870-180B-2
24	168	9.7	400	3	US-08-814-052-4
25	168	9.7	400	3	US-08-812-829-4
26	168	9.7	400	4	US-09-226-529-2
27	168	9.7	462	2	US-08-870-180B-13

28	168	9.7	462	4	US-09-226-529-13	Sequence 13, Appl
29	159	9.2	551	2	US-09-033-537A-1	Sequence 1, Appl
30	145.5	8.4	430	2	US-08-924-440-2	Sequence 2, Appl
31	143.5	8.3	956	4	US-09-134-078-63	Sequence 63, Appl
32	140	8.1	616	4	US-09-136-574A-47	Sequence 47, Appl
33	140	8.1	1426	4	US-09-136-574A-43	Sequence 43, Appl
34	139.5	8.1	327	2	US-08-169-948B-16	Sequence 16, Appl
35	139.5	8.1	327	2	US-08-448-873-16	Sequence 16, Appl
36	139.5	8.1	327	4	US-08-382-452D-16	Sequence 16, Appl
37	130.5	7.5	418	4	US-09-254-733-5	Sequence 5, Appl
38	124	7.2	613	4	US-09-149-727-5	Sequence 5, Appl
39	123.5	7.1	553	1	US-08-651-572-2	Sequence 2, Appl
40	123.5	7.1	553	3	US-09-066-544-2	Sequence 2, Appl
41	123.5	7.1	553	3	US-08-951-086-2	Sequence 2, Appl
42	123.5	7.1	553	4	US-09-430-669-2	Sequence 2, Appl
43	121	7.0	663	4	US-09-134-078-61	Sequence 61, Appl
44	121	7.0	680	4	US-09-134-078-25	Sequence 25, Appl
45	117.5	6.8	531	2	US-07-862-588B-7	Sequence 7, Appl

ALIGNMENTS

```
RESULT 1
US-09-066-075-2
; Sequence 2, Application US/09066075
; Patent No. 5925749
; GENERAL INFORMATION:
; APPLICANT: Mathur, E., et al.
; TITLE OF INVENTION: Carboxymethyl Cellulase from Thermotoga Maritima
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CARELLA, BYRNE, BAIN, GILFILLAN,
; STREET: 6 BECKER FARM ROAD
; CITY: ROSELAND
; STATE: NEW JERSEY
; COUNTRY: USA
; ZIP: 07068
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 INCH DISKETTE
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: WORD PERFECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/066,075
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/518,615
; FILING DATE: August 23, 1995
; ATTORNEY/AGENT INFORMATION:
; NAME: FERRARO, GREGORY D.
; REGISTRATION NUMBER: 36,134
; REFERENCE/DOCKET NUMBER: 331400-20
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-994-1700
; TELEFAX: 201-994-1744
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 317 AMINO ACIDS
; TYPE: AMINO ACID
; STRANDEDNESS:
; TOPOLOGY: LINEAR
; MOLECULE TYPE: PROTEIN
; US-09-066-075-2

Query Match 100.0% Score 1732; DB 2; Length 317;
Best Local Similarity 100.0%; Pred No. 6, 2e-161;
Matches 317; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 MGVDPERNKILRGINIGALAPNEGDMGVYIKDFPIIKKAGFSHVIRIPRSTHA 60
|||||
```

Db 1 MGVDPEFRNKILRGINIGNALPNEGDMGVVIKDEFFDIKEAGFSHVRIPIRMSTHA 60
QY 61 YAFPPYKIMDRFPKRVDEVINGALKRGLAVAINIHHEELMNDPEHKEFFLALMKQIAD 120
Db 61 YAFPPYKIMDRFPKRVDEVINGALKRGLAVAINIHHEELMNDPEHKEFFLALMKQIAD 120
QY 121 RYKDYPELFEILNEPFGNLTPEKWNELLEALKVIRSIDKHTIIGTAEMGGSIALE 180
Db 121 RYKDYPELFEILNEPFGNLTPEKWNELLEALKVIRSIDKHTIIGTAEMGGSIALE 180
QY 181 KLSVPMKWNKNSIVTIHYNNPEFTHOGAEWVGSEKMLGRKWSPPDQKHLIEEFNFTIE 240
Db 181 KLSVPMKWNKNSIVTIHYNNPEFTHOGAEWVGSEKMLGRKWSPPDQKHLIEEFNFTIE 240
QY 241 WSKKKRPIYIGEGFAYRKADLESRIKWTSPVVRMEKRRMSAWYEFCSGFGVYDTLRK 300
Db 241 WSKKKRPIYIGEGFAYRKADLESRIKWTSPVVRMEKRRMSAWYEFCSGFGVYDTLRK 300
QY 301 TWNKDLLEALIGDSTIE 317
Db 301 TWNKDLLEALIGDSTIE 317

RESULT 2
US-08-518-615A-2

; Sequence 2, Application US/08518615A
; Patent No. 5962258
; GENERAL INFORMATION:
; APPLICANT: Mathur, E., et al.
; TITLE OF INVENTION: Carboxymethyl Cellulase from Thermotoga Maritima
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CARELLA, BYRNE, BAIN, GILFILLAN,
; ADDRESS: CECCHI, STEWART & OLSTEIN
; STREET: 6 BECKER FARM ROAD
; CITY: ROSELAND
; STATE: NEW JERSEY
; COUNTRY: USA
; ZIP: 07068
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 INCH DISKETTE
; COMPUTER: IBM PS/2
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: WORD PERFECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/518,615A
; FILING DATE: August 23, 1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: FERRARO, GREGORY D.
; REGISTRATION NUMBER: 36,134
; REFERENCE/DOCKET NUMBER: 331400-20
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-994-1700
; TELEFAX: 201-994-1744
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 317 AMINO ACIDS
; TYPE: AMINO ACID
; STRANDEDNESS:
; TOPOLOGY: LINEAR
; MOLECULE TYPE: PROTEIN
; US-08-518-615A-2

Query Match 100.0%; Score 1732; DB 2; Length 317;
Best Local Similarity 100.0%; Pred. No. 6.2e-161;
Matches 317; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MGVDPEFRNKILRGINIGNALPNEGDMGVVIKDEFFDIKEAGFSHVRIPIRMSTHA 60
|||||

Db 1 MGVDPEFRNKILRGINIGNALPNEGDMGVVIKDEFFDIKEAGFSHVRIPIRMSTHA 60
QY 61 YAFPPYKIMDRFPKRVDEVINGALKRGLAVAINIHHEELMNDPEHKEFFLALMKQIAD 120
Db 61 YAFPPYKIMDRFPKRVDEVINGALKRGLAVAINIHHEELMNDPEHKEFFLALMKQIAD 120
QY 121 RYKDYPELFEILNEPFGNLTPEKWNELLEALKVIRSIDKHTIIGTAEMGGSIALE 180
Db 121 RYKDYPELFEILNEPFGNLTPEKWNELLEALKVIRSIDKHTIIGTAEMGGSIALE 180
QY 181 KLSVPMKWNKNSIVTIHYNNPEFTHOGAEWVGSEKMLGRKWSPPDQKHLIEEFNFTIE 240
Db 181 KLSVPMKWNKNSIVTIHYNNPEFTHOGAEWVGSEKMLGRKWSPPDQKHLIEEFNFTIE 240
QY 241 WSKKKRPIYIGEGFAYRKADLESRIKWTSPVVRMEKRRMSAWYEFCSGFGVYDTLRK 300
Db 241 WSKKKRPIYIGEGFAYRKADLESRIKWTSPVVRMEKRRMSAWYEFCSGFGVYDTLRK 300
QY 301 TWNKDLLEALIGDSTIE 317
Db 301 TWNKDLLEALIGDSTIE 317

RESULT 3

US-08-951-889-2
; Sequence 2, Application US/08951889
; Patent No. 6008032
; GENERAL INFORMATION:
; APPLICANT: Mathur, E., et al.
; TITLE OF INVENTION: Carboxymethyl Cellulase from
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CARELLA, BYRNE, BAIN, GILFILLAN,
; ADDRESS: CECCHI, STEWART & OLSTEIN
; STREET: 6 BECKER FARM ROAD
; CITY: ROSELAND
; STATE: NEW JERSEY
; COUNTRY: USA
; ZIP: 07068
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 INCH DISKETTE
; COMPUTER: IBM PS/2
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: WORD PERFECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/951,889
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: FERRARO, GREGORY D.
; REGISTRATION NUMBER: 36,134
; REFERENCE/DOCKET NUMBER: 331400-20
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-994-1700
; TELEFAX: 201-994-1744
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 317 AMINO ACIDS
; TYPE: AMINO ACID
; STRANDEDNESS:
; TOPOLOGY: LINEAR
; MOLECULE TYPE: PROTEIN
; US-08-951-889-2

Query Match 100.0%; Score 1732; DB 3; Length 317;
Best Local Similarity 100.0%; Pred. No. 6.2e-161;
Matches 317; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MGVDPEFRNKILRGINIGNALPNEGDMGVVIKDEFFDIKEAGFSHVRIPIRMSTHA 60
|||||


```

|||||
Db 1 MGVDPEFRNKILGRGINGNALPNEGDMGVYIKDEFFDILKEAGSHRIPIRMSTHA 60
OY 61 YAPPPYKIMDFEKKRVDEVINGALRKGLAVAINIHHEEELNDPEEKERFLAMKOIAD 120
Db 61 YAPPPYKIMDFEKKRVDEVINGALRKGLAVAINIHHEEELNDPEEKERFLAMKOIAD 120
OY 121 RYKDYPETLFEFELINEPHGNLTPEKNWELLEALKVIRSIDKHTTIIIGTAEMGISALE 180
Db 121 RYKDYPETLFEFELINEPHGNLTPEKNWELLEALKVIRSIDKHTTIIIGTAEMGISALE 180
OY 181 KLSVPKKEKNSIYTIHYNNPEFTHOGAEWVEGSEKWLGRKWSPPDQKHLIEEFNIEE 240
Db 181 KLSVPKKEKNSIYTIHYNNPEFTHOGAEWVEGSEKWLGRKWSPPDQKHLIEEFNIEE 240
OY 241 WSKKKNRPYIIGFEGAYRKADLESRIKWTSPVYREMEKRRMSNAWYBFCGFGVYDTLRK 300
Db 241 WSKKKNRPYIIGFEGAYRKADLESRIKWTSPVYREMEKRRMSNAWYBFCGFGVYDTLRK 300
OY 301 TWNKDLLEALIGDSIE 317
Db 301 TWNKDLLEALIGDSIE 317

```

RESULT 4
US-09-472-857-2
Sequence 2, Application US/09472857
Patent No. 6245547

```

GENERAL INFORMATION:
APPLICANT: Mathur, E., et al.
TITLE OF INVENTION: Carboxymethyl Cellulase from
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: CARELLA, BYRNE, BAIN, GILFILLAN,
STREET: 6 BECKER FARM ROAD
CITY: ROSELAND
STATE: NEW JERSEY
COUNTRY: USA
ZIP: 07068
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 INCH DISKETTE
COMPUTER: IBM PS/2
OPERATING SYSTEM: MS-DOS
SOFTWARE: WORD PERFECT 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/472,857
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/951,889
FILING DATE:
APPLICATION NUMBER: 08/518,615
FILING DATE: August 23, 1995
ATTORNEY/AGENT INFORMATION:
NAME: FERRARO, GREGORY D.
REGISTRATION NUMBER: 36,134
REFERENCE/DOCKET NUMBER: 331400-20
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201-994-1700
TELEFAX: 201-994-1744
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 317 AMINO ACIDS
TYPE: AMINO ACID
STRANDEDNESS:
TOPOLOGY: LINEAR
MOLECULE TYPE: PROTEIN
US-09-472-857-2

```

Query Match 100.0%; Score 1732; DB 4; Length 317;
Best Local Similarity 100.0%; Pred. No. 6,2e-161;

```

Matches 317; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 MGVDPEFRNKILGRGINGNALPNEGDMGVYIKDEFFDILKEAGSHRIPIRMSTHA 60
Db 1 MGVDPEFRNKILGRGINGNALPNEGDMGVYIKDEFFDILKEAGSHRIPIRMSTHA 60
OY 61 YAPPPYKIMDFEKKRVDEVINGALRKGLAVAINIHHEEELNDPEEKERFLAMKOIAD 120
Db 61 YAPPPYKIMDFEKKRVDEVINGALRKGLAVAINIHHEEELNDPEEKERFLAMKOIAD 120
OY 121 RYKDYPETLFEFELINEPHGNLTPEKNWELLEALKVIRSIDKHTTIIIGTAEMGISALE 180
Db 121 RYKDYPETLFEFELINEPHGNLTPEKNWELLEALKVIRSIDKHTTIIIGTAEMGISALE 180
OY 181 KLSVPKKEKNSIYTIHYNNPEFTHOGAEWVEGSEKWLGRKWSPPDQKHLIEEFNIEE 240
Db 181 KLSVPKKEKNSIYTIHYNNPEFTHOGAEWVEGSEKWLGRKWSPPDQKHLIEEFNIEE 240
OY 241 WSKKKNRPYIIGFEGAYRKADLESRIKWTSPVYREMEKRRMSNAWYBFCGFGVYDTLRK 300
Db 241 WSKKKNRPYIIGFEGAYRKADLESRIKWTSPVYREMEKRRMSNAWYBFCGFGVYDTLRK 300
OY 301 TWNKDLLEALIGDSIE 317
Db 301 TWNKDLLEALIGDSIE 317

```

RESULT 5
US-09-134-078-27
Sequence 27, Application US/09134078
Patent No. 6368844

```

GENERAL INFORMATION:
APPLICANT: Bylina, Edward J.
TITLE OF INVENTION: GLYCOSIDASE ENZYMES
NUMBER OF SEQUENCES: 72
CORRESPONDENCE ADDRESS:
ADDRESSEE: Gray Cary Ware & Freidenrich LLP
STREET: 4365 Executive Drive, Suite 1600
CITY: San Diego
STATE: CA
COUNTRY: USA
ZIP: 92121
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM compatible
OPERATING SYSTEM: Windows95
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/134,078
FILING DATE: 13-AUG-1998
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/949,026
FILING DATE: 10-OCT-1997
APPLICATION NUMBER: 60/056,916
FILING DATE: 06-DEC-1996
ATTORNEY/AGENT INFORMATION:
NAME: Hallie, Lisa A.
REGISTRATION NUMBER: 38,347
REFERENCE/DOCKET NUMBER: 09010/024002
TELECOMMUNICATION INFORMATION:
TELEPHONE: 858/677-1455
TELEFAX: 858/677-1456
INFORMATION FOR SEQ ID NO: 27:
SEQUENCE CHARACTERISTICS:
LENGTH: 360 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
FRAGMENT TYPE: internal
US-09-134-078-27

```

Query Match 48.2%; Score 835.5; DB 4; Length 360;


```

; APPLICANT: Tsai, Cheng-Fang
; APPLICANT: Chang, Chia-Chieh
; TITLE OF INVENTION: CELLULASE ENZYMES
; FILE REFERENCE: 08919/037001
; CURRENT APPLICATION NUMBER: US/09/419,459
; CURRENT FILING DATE: 1999-10-15
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 382
; TYPE: PRT
; ORGANISM: Piromyces rhizinflata
US-09-419-459-2

```

```

Query Match          17.8%: Score 308; DB 4; Length 582;
Best Local Similarity 26.3%: Pred. No. 1, 1e-21;
Matches 83; Conservative 54; Mismatches 127; Indels 52; Gaps 8;

```

```

QY 10 KILGGINIGALEA-----PNEGDMG-VVTKDEFEDIKAGFSHVRIPIR 55
DB 148 KEMNFGMNLGNTMDACIEYLVEKDOTASETCWGNPKTTEMFKVLIDNCFVFRIPPT 207
QY 56 WSTHVAAPRYKIMDFEPRKRVDEVINGALRGALVAIINH-----YEELMNDPEHKE 109
DB 208 WSGHGEADPYKIDEXMKRVHEVVDYPRKNGAFVILNLHETMNAFSETLDTAKEILE 267
QY 110 RFLAMKQIADRYKDYPTLFFELINEPHGNLTPEKW-----NELLEALKVIR 158
DB 268 K---IMSQIAEERKDYDEHLIEGLNEPRKNDTPVEMTGDDGDMADVAMNMFUKTVR 324
QY 159 SI---DKHTTIIIGTAEWGGSISALEKLSVPMKKNISYVTHIYNPFEFTHOGAEWVEGS 214
DB 325 SAGGNPKKHLMPYAAACNENSNFNFTPEDDDVIVASVHAAYNPNLNNBEGADVCK 384
QY 215 EKWIGRK---WSPDDOKHLEEFNIEEMSKNKRPIYIGEGAYRKADLESRIKWTSF 271
DB 385 FIDAGRKDELMNLMKKRRVDD-----GIPMLIGEGAMNRDNEEDRATVAER 433
QY 272 VVREKRRKRWMAWYE 287
DB 434 YMEKVTAMGVPOIWM 449

```

```

RESULT 12
US-08-732-433-1
; Sequence 1, Application US/08732433
; Patent No. 6063611
; GENERAL INFORMATION:
; APPLICANT: van Solingen, Dieter
; TITLE OF INVENTION: No. 6063611el Alkaline Cellulase and
; NUMBER OF INVENTION: Method of Producing Same
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genencor International, Inc.
; STREET: 925 Page Mill Road
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304-1013
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/732,433
; FILING DATE: 18-APR-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US96/05651
; FILING DATE: 26-APR-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Stone, Christopher L
; REGISTRATION NUMBER: 35,696

```

```

; REFERENCE/DOCKET NUMBER: GC287-US2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-846-7620
; TELEFAX: 650-845-6504
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 574 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-732-433-1

```

```

Query Match          17.7%: Score 306.5; DB 3; Length 574;
Best Local Similarity 26.0%: Pred. No. 1, 5e-21;
Matches 87; Conservative 57; Mismatches 126; Indels 65; Gaps 12;

```

```

QY 15 GINIGNALEA--PNEGDMG-VVTKDEFEDIKAGFSHVRIPIRWSHVAAPRYKIMOR 71
DB 45 GNNLGTEDAVGQDETANGNPRVTRELIERTADGYSIRIPVTWENRIGGADYPIDPO 104
QY 72 FPKRYDEVINGALRGALVAIINHIEELMNDPEHKE-----ERFLAMKQIADRYKDYR 126
DB 105 FLNRVDEVVQWLEEDLYVMNLHDSWLVMTYEMHNNGVMKXRYSLMQLSHHEKDYR 164
QY 127 ETLPFEELINEPHGNLTPEKMWELLE--EALKVIRSIDKHTIIG--TAEWGGSISALEKLS 183
DB 165 TKLMFESVNERPFS---QWGEIRRNHAL-----LDDLNTVFPEIYROSOGONDIRPLV 216
QY 184 VKRWE-----KNSIVTHIYNPFEFTHOGAEWVEGSEKKLGRKWG 223
DB 217 LPTMETATSQPLNMLYOTIDKLDPPNLAIVTHYIGFVPSVNIAGITREE----- 268
QY 224 SPDDOKHLEEFNIEEMSKNKRPIYIGERG--AYRK---ADLESRIKWTSVVBE 277
DB 269 --DSKREIETDRVHNHTVANGIPVLAEGEGLGFDKHTGVIOGGEKLFKEFYLIHLN 326
QY 278 KRRSWAWYEFCSGEG-----VYDTLAKTY 302
DB 327 ERDITHLMDNGQHFRNHTYEWYDEELDMLRASW 361

```

```

RESULT 13
US-08-945-574-2
; Sequence 2, Application US/08945574
; Patent No. 6313081
; GENERAL INFORMATION:
; APPLICANT: Ienting, Hermanus Bernardus Maria
; APPLICANT: Maurel, Karl-Heinz
; APPLICANT: Kottwitz, Beatrix
; APPLICANT: Weiss, Albrecht
; APPLICANT: van Solingen, Pieter
; TITLE OF INVENTION: Detergents Comprising Cellulases
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Henkel Corporation
; STREET: 140 Germantown Pike, Suite 150
; CITY: Plymouth Meeting
; STATE: Pennsylvania
; COUNTRY: U.S.A.
; ZIP: 19462
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" diskette
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: MS Word 6.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/945,574
; FILING DATE: unavailable
; CLASSIFICATION:
; CLASSIFICATION: C11D 3/386
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/EP96/01755

```

FILING DATE: 26 Apr. 1996
 APPLICATION NUMBER: EP 95201115.3
 FILING DATE: 28 Apr. 1995
 APPLICATION NUMBER: U.S. 614.115
 FILING DATE: 12 Mar. 1996
 ATTORNEY/AGENT INFORMATION:
 NAME: Murphy, Glenn E. J.
 REGISTRATION NUMBER: 33,539
 REFERENCE/DOCKET NUMBER: H 1920 PCT/US
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (610) 832-2228
 TELEFAX: (610) 941-6067
 INFORMATION FOR SEQ ID NO: 2:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 574 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-08-945-574-2

Query Match 17.7% Score 306.5; DB 4; Length 574;
 Best Local Similarity 26.0%; Pred. No. 1.5e-21;
 Matches 87; Conservative 57; Mismatches 126; Indels 65; Gaps 12;

QY 15 GINIGNALAE--PNEGDMG-VVIRDEFFDIKEAGFSHVRIPIRIMSTHAYAFPPYKIMDR 71
 DB 45 GWNIGNFEDAVGOETAMGNPRVRELIERADGYSIRIPYWEHRIGAPYIDPQ 104
 QY 72 FFKVDEVINGALKRGALVAIINHIEELMNDPEERK---EFLALMKOIAADRYDYP 126
 DB 105 FLNKDEVYQWALBEDLYVMIINHDSMLYEMEHNYNGMAKYRSIMEDLSMHPKDY 164
 QY 127 ELTFEELINEPHGNLTPEKNNELLE--EALKVIRSIDKHHITIG-TAEMGIGISALKLS 183
 DB 165 TKLMESVNEPKES---QNMGEIRENNHAL-----LDDLNVFEFYIRQSGGQNDIRPLV 216
 QY 184 VPKWE-----KNSIVTIHYNPEEFTHOGAEVWEGSEKMLGRKWG 223
 DB 217 LPTWETATQPLNLNLYQITIDKLDPLNLIATVHYHGFWSPSVNAGTTRFEE----- 268
 QY 224 SPDOQKILIEFNFIEMSKKNRPYIYIGFG--AYRK---ADLESRIKWTSVVAREME 277
 DB 269 --DSKRRIIEFFDVRHHTFVARGIPVVLGEGFLGFDKHTGVIOQGRKLFEEYLIIHLN 326
 QY 278 KRMSMAWYEFCSGFG-----YVDPLRKTW 302
 DB 327 ERDITHMLMDNGQHFNRHTTEWIDEELFDMLRASW 361

RESULT 14
 US-09-118-324-5
 Sequence 5, Application US/09118324
 Patent No. 6110720
 GENERAL INFORMATION:
 APPLICANT: Li, Xin-Liang
 APPLICANT: Ljungdahl, Lars G.
 APPLICANT: Chen, Huizhong
 TITLE OF INVENTION: Orpinomyces Cellulase Celc Protein and Coding Sequences
 FILE REFERENCE: 32-96Sequence listing
 CURRENT APPLICATION NUMBER: US/09/118.324
 NUMBER OF SEQ ID NOS: 5
 SOFTWARE: Patentln Ver. 2.0
 SEQ ID NO 5
 LENGTH: 471
 TYPE: PRT
 ORGANISM: Orpinomyces sp. PC-2
 US-09-118-324-5

Query Match 17.1% Score 296.5; DB 3; Length 471;
 Best Local Similarity 28.0%; Pred. No. 1e-20;
 Matches 98; Conservative 51; Mismatches 134; Indels 67; Gaps 14;

QY 10 KILRGINIGNALAE-----PNEGDMG-VVIRDEFFDIKEAGFSHVRIPIR 55
 DB 30 KELTIGSLGNTLDASCVETLNTSKDQASCTCGNKKTOELYYKLSDLGCFNFRIPPT 89
 QY 56 WSTHAYAFPPYKIMDRFFKRVDEVINGALKRGALVAIINHIEELMNDPEERK 110
 DB 90 WSGHFGDAPDYKIDSVMMKRVHEVVDYALNTGCAIILNIH--ETWVYAFQKNLESARKI 147
 QY 111 FLALMKOIAADRYKDYPELTFEELINEPHGNLTPEPKM-----NELLEALKYRS 159
 DB 148 LVAIWKQIAEFQYDEHLTFEGNNEPRKYVDPAEWGTGGDEGNFVEMNALFVKITRA 207
 QY 160 ID-----KKHTIIIGTAEMGISALEKLSVPKMEKNSIVTIHYNPEFT--HOG-----AEW 210
 DB 208 TGGNANRHLIMPIYASVNDGSIINNKRYPRGDDKIVLSHSYSPYFALNNGGATSNF 267
 QY 211 VEGSE-KWLGKRWGSPDDQKHLIEFNFIEEMSKKNRPYIYIGFGAYRKADLESRIKWT 269
 DB 268 YDGNIEIDMV-----MNTINSSFTSKGIPVITIGEFVAMNRDNEDDRERWQ 311
 QY 270 SFVYRENEKRMSWAYE--FCSG---FGYVDPLRKTNMDLEALIG 313
 DB 312 EYTIKATATGIPCVIWDNGYFEGGERFGIID--RKSINV-IEPKLNG 358

RESULT 15
 US-09-286-691-12
 Sequence 12, Application US/09286691
 Patent No. 6190189
 GENERAL INFORMATION:
 APPLICANT: Li, Xin-Liang
 APPLICANT: Ljungdahl, Lars G.
 APPLICANT: Chen, Huizhong
 TITLE OF INVENTION: Cellulases and Coding Sequences
 FILE REFERENCE: 42-96
 CURRENT APPLICATION NUMBER: US/09/286.691
 EARLIER FILING DATE: 1999-04-05
 EARLIER APPLICATION NUMBER: US 60/027.883
 EARLIER FILING DATE: 1996-10-04
 EARLIER APPLICATION NUMBER: PCT US97/18008
 EARLIER FILING DATE: 1997-10-03
 NUMBER OF SEQ ID NOS: 29
 SOFTWARE: Patentln Ver. 2.0
 SEQ ID NO 12
 LENGTH: 471
 TYPE: PRT
 ORGANISM: Orpinomyces sp. PC-2
 US-09-286-691-12

Query Match 17.1% Score 296.5; DB 4; Length 471;
 Best Local Similarity 28.0%; Pred. No. 1e-20;
 Matches 98; Conservative 51; Mismatches 134; Indels 67; Gaps 14;

QY 10 KILRGINIGNALAE-----PNEGDMG-VVIRDEFFDIKEAGFSHVRIPIR 55
 DB 30 KELTIGSLGNTLDASCVETLNTSKDQASCTCGNKKTOELYYKLSDLGCFNFRIPPT 89
 QY 56 WSTHAYAFPPYKIMDRFFKRVDEVINGALKRGALVAIINHIEELMNDPEERK 110
 DB 90 WSGHFGDAPDYKIDSVMMKRVHEVVDYALNTGCAIILNIH--ETWVYAFQKNLESARKI 147
 QY 111 FLALMKOIAADRYKDYPELTFEELINEPHGNLTPEPKM-----NELLEALKYRS 159
 DB 148 LVAIWKQIAEFQYDEHLTFEGNNEPRKYVDPAEWGTGGDEGNFVEMNALFVKITRA 207
 QY 160 ID-----KKHTIIIGTAEMGISALEKLSVPKMEKNSIVTIHYNPEFT--HOG-----AEW 210
 DB 208 TGGNANRHLIMPIYASVNDGSIINNKRYPRGDDKIVLSHSYSPYFALNNGGATSNF 267
 QY 211 VEGSE-KWLGKRWGSPDDQKHLIEFNFIEEMSKKNRPYIYIGFGAYRKADLESRIKWT 269
 DB 268 YDGNIEIDMV-----MNTINSSFTSKGIPVITIGEFVAMNRDNEDDRERWQ 311

Mon Apr 28 09:32:03 2003

us-09-880-729a-2.raii

Page 8

```

Qy      270 SPVAREMERKRRSMAYNE--FCSG---EGYDTERKTKNKDLEALITG 313
      : ::: |::: |::: |::: |::: |::: |::: |::: |::: |
Db      312 EYVKKATAGLIPCVMIDNGEYFEGEGEGERGID--RKSLNV--IFPKLNG 358

```

Search completed: April 22, 2003, 14:01:19
Job time : 32 secs

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 22, 2003, 13:59:59 ; Search time 20 Seconds
(without alignments)
1270.057 Million cell updates/sec

Title: US-09-880-729A-2

Perfect score: 1732

Sequence: 1 MGVDFERRKILGRGINICN.....LRTKWNKLLLEALIGDSIE 317

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 301932 seqs, 80129803 residues

Total number of hits satisfying chosen parameters: 301932

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications_AA:*

- 1: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep:*
- 2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep:*
- 3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep:*
- 4: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep:*
- 5: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:*
- 6: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:*
- 7: /cgn2_6/ptodata/1/pubpaa/PCT07_PUBCOMB.pep:*
- 8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep:*
- 9: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep:*
- 10: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep:*
- 11: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep:*
- 12: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB.pep:*
- 13: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep:*
- 14: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1732	100.0	317	9	US-09-880-729-2
2	835.5	48.2	360	9	US-10-121-032-27
3	306.5	17.7	574	10	US-09-863-547B-2
4	292	16.9	395	9	US-09-797-464A-2
5	285	16.5	400	9	US-09-797-464A-4
6	283.5	16.4	348	9	US-09-797-464A-7
7	279	16.1	363	9	US-09-797-464A-11
8	203	11.7	160	9	US-09-797-464A-8
9	203	11.7	167	9	US-09-797-464A-6
10	177	10.2	308	9	US-10-068-374-5
11	177	10.2	467	9	US-10-068-374-2
12	177	10.2	467	10	US-09-863-547B-1
13	168	9.7	400	9	US-09-903-185-2
14	168	9.7	462	9	US-09-903-185-13
15	159.5	9.2	335	9	US-09-927-827-52
16	147.5	8.5	1621	9	US-10-185-990-10
17	143.5	8.3	956	9	US-10-121-032-63
18	139.5	8.1	327	9	US-09-916-494A-16
19	138.5	8.0	333	9	US-09-927-827-51

20	136.5	7.9	535	9	US-09-927-827-50	Sequence 50, Appl
21	129	7.4	325	9	US-09-927-827-17	Sequence 17, Appl
22	128	7.4	427	9	US-09-850-982B-2	Sequence 2, Appl
23	123.5	7.1	553	10	US-09-888-224-2	Sequence 2, Appl
24	121.5	7.0	558	9	US-09-862-660-2	Sequence 2, Appl
25	121	7.0	663	9	US-10-121-032-61	Sequence 61, Appl
26	121	7.0	680	9	US-10-121-032-25	Sequence 25, Appl
27	116.5	6.7	454	9	US-10-121-032-20	Sequence 20, Appl
28	112.5	6.5	166	9	US-09-797-464A-5	Sequence 5, Appl
29	111.5	6.4	174	9	US-09-797-464A-9	Sequence 9, Appl
30	110.5	6.4	333	9	US-09-927-827-3	Sequence 3, Appl
31	108	6.2	562	10	US-09-981-900B-5	Sequence 5, Appl
32	105	6.1	464	10	US-09-815-242-10647	Sequence 10647, A
33	105	6.1	467	10	US-09-815-242-4997	Sequence 4997, Ap
34	96.5	5.6	423	10	US-09-369-735B-5	Sequence 2, Appl
35	96.5	5.6	423	10	US-09-369-735B-5	Sequence 5, Appl
36	95.5	5.5	329	9	US-09-927-827-16	Sequence 16, Appl
37	95	5.5	602	9	US-10-195-158-5	Sequence 5, Appl
38	93	5.4	421	9	US-10-121-032-17	Sequence 17, Appl
39	93	5.4	511	9	US-10-121-032-21	Sequence 21, Appl
40	92.5	5.3	515	10	US-09-765-272-22	Sequence 22, Appl
41	92.5	5.3	684	9	US-10-121-032-62	Sequence 62, Appl
42	91.5	5.3	459	10	US-09-815-242-13229	Sequence 13229, A
43	90.5	5.2	370	9	US-09-738-626-5196	Sequence 5196, Ap
44	90	5.2	489	10	US-09-369-735B-9	Sequence 9, Appl
45	87	5.0	438	9	US-10-121-032-16	Sequence 16, Appl

ALIGNMENTS

RESULT 1
US-09-880-729-2
Sequence 2, Application US/09880729
Publication No. US20030044956A1
GENERAL INFORMATION:
APPLICANT: Matchur, E., et al.
TITLE OF INVENTION: Carboxymethyl Cellulase from
Thermotoga Maritima
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: CARELLA, BYRNE, BAIN, GILFILLAN,
CECCHI, STEWART & OLSTEIN
STREET: 6 BECKER FARM ROAD
CITY: ROSELAND
STATE: NEW JERSEY
COUNTRY: USA
ZIP: 07068
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 INCH DISKETTE
COMPUTER: IBM PS/2
OPERATING SYSTEM: MS-DOS
SOFTWARE: WORD PERFECT 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/880,729
FILING DATE: 12-Jun-2001
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/09/472,857
FILING DATE: <Unknown>
APPLICATION NUMBER: 08/951,889
FILING DATE: <Unknown>
APPLICATION NUMBER: 08/518,615
FILING DATE: August 23, 1995
ATTORNEY/AGENT INFORMATION:
NAME: FERRARO, GREGORY D.
REGISTRATION NUMBER: 36,134
REFERENCE/DOCKET NUMBER: 331400-20
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201-994-1700
TELEFAX: 201-994-1744
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:

LENGTH: 317 AMINO ACIDS
TYPE: AMINO ACID
STRANDEDNESS: <Unknown>
TOPOLOGY: LINEAR
MOLECULE TYPE: PROTEIN
SEQUENCE DESCRIPTION: SEQ ID NO: 2:
us-09-880-729-2

Query Match 100.0%; Score 1732; DB 9; Length 317;
Best Local Similarity 100.0%; Pred. No. 3,2e-146;
Matches 317; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MGVDPERRKILIGRGINIGNALAPNEGDMGVYIKDEFDIIKEAGFSVRIPIRSTHIA 60
DB 1 MGVDPERRKILIGRGINIGNALAPNEGDMGVYIKDEFDIIKEAGFSVRIPIRSTHIA 60
QY 61 YAFPPYKIDREFKRYDEVINGALKRGLAVAINIHHEELMNDPEEKKERFLALMQIAD 120
DB 61 YAFPPYKIDREFKRYDEVINGALKRGLAVAINIHHEELMNDPEEKKERFLALMQIAD 120
QY 121 RYKDYPELTFEELNPHGNLTPEKWNELLEALKVIRSIDKKHTIIIGTAEMGGSIALE 180
DB 121 RYKDYPELTFEELNPHGNLTPEKWNELLEALKVIRSIDKKHTIIIGTAEMGGSIALE 180
QY 181 KISVPKWEKNSIVTIIHYNPPEFTHOGARWEGSEKMLGRKMGSPDDOKHLEEFNFIEE 240
DB 181 KISVPKWEKNSIVTIIHYNPPEFTHOGARWEGSEKMLGRKMGSPDDOKHLEEFNFIEE 240
QY 241 WSKNNRPYIYIGFGAYRRADLESRIKWTSEVVREREMKRRMSAWYEFCSGFGVDTLTK 300
DB 241 WSKNNRPYIYIGFGAYRRADLESRIKWTSEVVREREMKRRMSAWYEFCSGFGVDTLTK 300
QY 301 TWNKDLLEALLIGDSTE 317
DB 301 TWNKDLLEALLIGDSTE 317

RESULT 2
Sequence 27, Application US/10121032
Patent No. US2002015550A1
GENERAL INFORMATION:
APPLICANT: Bylina, Edward J.
TITLE OF INVENTION: GLYCOSIDASE ENZYMES
NUMBER OF SEQUENCES: 72
CORRESPONDENCE ADDRESS:
ADDRESSEE: Gray Cary Ware & Freidenrich LLP
STREET: 4365 Executive Drive, Suite 1600
CITY: San Diego
STATE: CA
COUNTRY: USA
ZIP: 92121
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: Windows95
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/121,032
FILING DATE: 09-Apr-2002
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/09/134,078
FILING DATE: 13-AUG-1998
APPLICATION NUMBER: 08/949,026
FILING DATE: 10-OCT-1997
APPLICATION NUMBER: 60/056,916
FILING DATE: 06-DEC-1996
ATTORNEY/AGENT INFORMATION:
NAME: Hallie, Lisa A.
REGISTRATION NUMBER: 38,347
REFERENCE/DOCKET NUMBER: 09010/024002
TELECOMMUNICATION INFORMATION:

TELEPHONE: 858/677-1456
TELEFAX: 858/677-1465
INFORMATION FOR SEQ ID NO: 27:
SEQUENCE CHARACTERISTICS:
LENGTH: 360 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
FRAGMENT TYPE: internal
SEQUENCE DESCRIPTION: SEQ ID NO: 27:
us-10-121-032-27

Query Match 48.2%; Score 835.5; DB 9; Length 360;
Best Local Similarity 49.5%; Pred. No. 2e-66;
Matches 154; Conservative 53; Mismatches 95; Indels 9; Gaps 3;

QY 6 FERNKILGRGINIGNALAPNEGDMGVYIKDEFDIIKEAGFSVRIPIRSTHIAVFP 65
DB 51 FEYKMWGKGVNIGNALAPNEGDMGVRIEDEYELIKRGFDSVRIPIRSTHIAISKPP 110
QY 66 YKIMDFEFKRYDEVINGALKRGLAVAINIHHEELMNDPEEKKERFLALMQIADRYKDY 125
DB 111 YDIDRNFLEKRVHVVDRALENNLTIVINTHFEELYQEPDKGVYVLEIMRQIAKFEKDY 170
QY 126 PETLEFELNPHGNLTPEKWNELLEALKVIRSIDKKHTIIIGTAEMGGSIALEKISVP 185
DB 171 PETLEFELNPHGNLTPEKWNELLEALKVIRSIDKKHTIIIGTAEMGGSIALEKISVP 185
QY 186 KMEKNSIVTIIHYNPPEFTHOGARWEGSEKMLGRKMGSPDDOKHLEEFNFIEE 241
DB 231 N-DKRIIVSFHYEPEKRYHOGAPVNPPIPVYKVNCEWEIINDINSH---FKYSDW 285
QY 242 SKNNRPYIYIGFGAYRRADLESRIKWTSEVVREREMKRRMSAWYEFCSGFGVDTLTK 301
DB 286 AKONNYPPIFLGFGAYRRADLESRIKWTSEVVREREMKRRMSAWYEFCSGFGVDTLTK 301
QY 302 WNKDLLEALLIG 312
DB 346 WIEPLATAVVG 356

RESULT 3
US-09-863-547B-2
Sequence 2, Application US/09863547B
Patent No. US20020128166A1
GENERAL INFORMATION:
APPLICANT: Henkel KGaA
APPLICANT: Hermanns, Lentus B. M.
APPLICANT: Van Beckhoven, Rudolf F. W. C.
APPLICANT: Maurer, Karl-Heinz
APPLICANT: Kottwitz, Beatrix
APPLICANT: Weis, Albrecht
TITLE OF INVENTION: Detergents Comprising Cellulases
FILE REFERENCE: H 1920 A
CURRENT APPLICATION NUMBER: US/09/863,547B
CURRENT FILING DATE: 2002-02-13
PRIOR APPLICATION NUMBER: US 08/945,574
PRIOR FILING DATE: 1998-02-27
PRIOR APPLICATION NUMBER: PCT/EP96/01755
PRIOR FILING DATE: 1996-04-26
PRIOR FILING DATE: 1996-03-12
PRIOR APPLICATION NUMBER: US 08/614,115
PRIOR FILING DATE: 1995-04-28
NUMBER OF SEQ ID NOS: 2
SOFTWARE: PatentIn version 3.1
SEQ ID NO 2
LENGTH: 574
TYPE: PRT
ORGANISM: Bacillus sp. 669.93
US-09-863-547B-2


```

: TITLE OF INVENTION: Family 5 xyloglucanases
:
: FILE REFERENCE: 6073.200-US
:
: CURRENT APPLICATION NUMBER: US/09/797,464A
:
: CURRENT FILING DATE: 2002-02-19
:
: NUMBER OF SEQ ID NOS: 25
:
: SOFTWARE: PatentIn version 3.1
:
: SEQ ID NO: 7
:
: LENGTH: 348
:
: TYPE: prt
:
: ORGANISM: paenibacillus pabuli
:
: US-09-797-464A-7

```

Query Match	16.4%	Score 283.5	DB 9	Length 348
Best Local Similarity	25.4%	Pred No. 2.3e-17		
Matches 86, Conservative	56	Mismatches 117	Indels 79	Gaps 13

```

QY      12 LGRGINGNMALEA-----PNCMDG-VIDEDEFDIDKEAGFSHVRPIIWSHTAAVFP 65
        1 : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      2 MGAGNLDNQLAALVANGIPNBTAMGNFTYPIBELIKKAKGAFSIRIPVSLNNISASN 61
QY      66 YKIMDFRKRDEYINGALKRGALVAIINH--HYE-----LMDPEEH--KERLALM 115
        1 : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      62 YTNAMAMNRIQQVVDVAYVNGELVYIINIHGDGNSVQGMGLVNGNGQTAIREKKYKW 121
QY      116 KOIADRYKDPETTEFFELNE-----PIGNHLPKEMNELLEALVKIRISDKHTIIIGTA 171
        : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      122 QOIATKESNYNDHLIEESMNEVPOGNTGNPSAY-----ANLANYQIIVVDIV 170
QY      172 -EMGISALEKLSVPKW-----EKNSTVTHIYNNPE 202
        : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      171 ROTGGNNNAKWLIPGWNTINIDFTVGNYGFPALPTDNFRSSAIPBSQKRIKISAHYSFMD 230
QY      203 F-----THOGAEWVGESEKWLGRKKGSPDDOKHLEENFIETWSKNNRPYIGEE 254
        : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      231 FAGEENGNTQMGATATNPARK-----STWGQED---YLESQKSMYKDFVQGVPIGEE 284
QY      255 GAYRKADLES-----RIKWTFSYVREMEKRRMSAYNE 287
        1 : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      285 GSIDKTSYDSTNNVYRAAYAAVATAKAKKIYMAVPIYMD 322

```

```

RESULT 7
US-09-797-464A-11
: Sequence 11, Application US/09797464A
: Publication No. US2003002807A1
: GENERAL INFORMATION:
: APPLICANT: Wilting, Reinhard
: APPLICANT: Bjornqvad, Mats Eskelund
: APPLICANT: Kauppinen, Markus Sakari
: APPLICANT: Schulten, Martin
: TITLE OF INVENTION: Family 5 xyloglucanase
: FILE REFERENCE: 6073.200-US
: CURRENT APPLICATION NUMBER: US/09/797,464A
: CURRENT FILING DATE: 2002-02-19
: NUMBER OF SEQ ID NOS: 25
: SOFTWARE: PatentIn version 3.1
: SEQ ID NO 11
: LENGTH: 163
: TYPE: PRT
: ORKANSMA: Paenibacillus pabuli
US-09-797-464A-11

```

Query Match	16.1%	Score 279;	DB 9;	Length 363;	.
Similarity	24.1%	Pred. No. 6	1e-17;		
Best Local					
Matches 89;	Conservative	66;	Mismatches 130;	Indels 84;	Gaps 15

[illegible]

```

QY 116 KQJADRXKXYPPELEFFELLNE-----PGNLTPEKXNMLEEALKVIRSDKKHHIIIGTA 117
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 130 QOQVATFESNTNDLTFESMNEYFDGNGNPNATAY-----ANLNVNNDLFDVJ 178
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 172 -EMOGISALEKTSVPMK-----EKXSIVTHIYWEFE 202
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 179 ROTGCGNNNAWMLITGOMNTNIDFTVGNYGALPTDNRSSAIPSSQKRIMISAHYSPMD 238
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 203 F-----THOGAEWVEGSEKWKJGRKQSPDDQKHLIEEFNFIETWSAKKKRPYITGEF 254
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 239 FAGEENGNTTQMGATYTNPAKK---STMQGD---YLESQFSKMDKFPYTGQYPVVIGE 292
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 255 GAYKRADLES-----RIKWTFSVVEHEKRRWSNAIWE-----FCSGFGVYDTLTKT *WK 304
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 293 GSDIKTSTYDSNNVYFAAALAKAVAKAKKRYKKMVPYWDGSHGNGHGFALFNKQNNVTYQO 352
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 305 DLLEALIG 313
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 353 MNVNAIMQG 361
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

```

RESULT 8
US-09-797-464A-8
Sequence 8, Application US/09797464A
Publication No. US20030022807A1
GENERAL INFORMATION:
APPLICANT: Wilting, Reinhard
APPLICANT: Bjornqvist, Mats Eskelund
APPLICANT: Kauppien, Markus Sakari
APPLICANT: Schulten, Martin
TITLE OF INVENTION: Family 5 Xylolucanase
FILE REFERENCE: 6073.200-US
CURRENT APPLICATION NUMBER: US/09/797,464A
CURRENT FILING DATE: 2002-02-19
NUMBER OF SEQ ID NOS: 25
SOFTWARE: PatentIn version 3.1
SEQ ID NO 8
LENGTH: 160
TYPE: prt
ORGANISM: Paenibacillus pabuli
US-09-797-464A-8

```

Query Match	11.7%	Score 203;	DB 9;	Length 160;
Best Local Similarity	32.7%	Pred. No. 1.2e-10;		
Matches	52;	Conservative	25;	Mismatches 54;
			Indels	28;
			Gaps	6;

QY 13 GKGINIGNALLA-----PREGSWG-VVIKDEFDILIEAGSHRIPDIRMSTHAYAPPY 66
 | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 1 GAGWNIIGNLEAAVNGTPTNETATMGNPPTYPELIIKKVKAAGEKSRITREYSTUNINGSAFNT 60

QY 67 KIMDRERFRVDENVINGALKRGILAVALINTH--HYEE-----LMMDPEEH--KERFLALMK 116
 | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 61 TTAAMALNLIOGVADVAVANEGLYLIITHIDGYSVCGGLLVANGNOTAIKEKYKRWQ 120

QY 117 QIADRYKDYPETLFEELINEP-----HGNTLP 143
 | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 121 QIATKFSNNDRIEESMAVEFDGYCNPNSAYIANLNP 159

```

RESULT 9
US-09-797-464A-A-6
; Sequence 6, Application US/09797464A
; Publication No. US20030022807A1
; GENERAL INFORMATION:
; APPLICANT: Wiltung, Reinhard
; APPLICANT: Bjornvad, Mads Eskelund
; APPLICANT: Kauppinen, Markus Sakari
; APPLICANT: Schulten, Martin
; TITLE OF INVENTION: Family 5 Xyloglucanases
; FILE REFERENCE: 6073.200-US
; CURRENT APPLICATION NUMBER: US/09/797,464A
; CURRENT FILING DATE: 2002-02-19

```

```
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 6
; LENGTH: 167
; TYPE: PRF
; ORGANISM: Paenibacillus pabuli
US-09-797-464A-6
```

```
Query Match          11.7%; Score 203; DB 9; Length 167;
Best Local Similarity 35.0%; Pred. No. 1.3e-10;
Matches 49; Conservative 25; Mismatches 50; Indels 16; Gaps 5;
```

```
QY 13 GRCIGNMALEA-----PNEGDMG-VVVKDEFDIIKEAGFSHVRIPIRSTHAYAPPY 66
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1 GAGNMLGNLEAAVNGPTNPTAMGNPTVPELIIKKVAKGFSIRIPVSYLSNIGSAPNY 60
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 67 KIDRFKRDVEYNGKLKGLAVAINIH--HFE-----LMDPEEH--KERFLAMK 116
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 61 TINAAWLNRIQQVVDYVYNEGVLVINIHGDYNSVQGMVLVNGNQTAKKKKKWQ 120
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 117 QIADRYKVPETLEFFELINE 136
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 121 QIATKFSNYNDRLIFESMNE 140
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
```

```
RESULT 10
US-10-068-374-5
; Sequence 5, Application US/10068374
; Publication No. US20030044903A1
; GENERAL INFORMATION:
; APPLICANT: Shaw, Andrew
; TITLE OF INVENTION: Modified Target Enzymes Having Catalytic
; TITLE OF INVENTION: Triads
; TITLE OF INVENTION: and Uses Thereof
; FILE REFERENCE: GC652
; CURRENT APPLICATION NUMBER: US/10/068,374
; CURRENT FILING DATE: 2002-02-06
; PRIOR APPLICATION NUMBER: US 60/268,347
; PRIOR FILING DATE: 2001-02-12
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 308
; TYPE: PRF
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: protein encoded by plasmid pCORE3
US-10-068-374-5
```

```
Query Match          10.2%; Score 177; DB 9; Length 308;
Best Local Similarity 25.3%; Pred. No. 5.8e-08;
Matches 67; Conservative 40; Mismatches 94; Indels 64; Gaps 15;
```

```
QY 29 DMGVVVKDEFDIIKEAGFSHVRIPIRSTHAYAPPYKIMDRFKRVDEVINGALKRGL 88
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 55 DMGIVT-----FRAMYTSSGGYIDP-----SVKEKYKETEVEAIDLGI 94
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 89 AVAINIHHELMNDPEEHKERFLAMKQIADRYKVPETLEFFELINEPHGNLPEKMWNE 148
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 95 YVILDMHILSD--NDPVIYKEAKDFDEMSLXGDPNVI-YEIAEPNG--SDVTWDN 149
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 149 LL-----EALAKVIRSIDKKHTIIGTAEWG---GISALEKLSVPKWKNSIVTTHYNNP 201
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 150 QIKRYAEVIVPIRDNDPNIVIGTWSQDVHHAADNQLADP-----NWYAFHFY--- 202
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 202 EFTHQGAWEWEGSEKWKGRKWSGPDOKHLIEFNFLIEMSKKKRRPIYIGEGAYRKAD 261
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 203 AGTH-GGNLRDQVYAL-----DGAAL-----FVSEWGTG---AATGDGGVF----- 241
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 262 LESRIKWTFSVVREREMKRRMSWAYW 286
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 242 LDEAQVWIDF-----MDERNLSMANW 262
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
```

```
RESULT 11
US-10-068-374-2
; Sequence 2, Application US/10068374
; Publication No. US20030044903A1
; GENERAL INFORMATION:
; APPLICANT: Shaw, Andrew
; TITLE OF INVENTION: Modified Target Enzymes Having Catalytic
; TITLE OF INVENTION: Triads
; TITLE OF INVENTION: and Uses Thereof
; FILE REFERENCE: GC652
; CURRENT APPLICATION NUMBER: US/10/068,374
; CURRENT FILING DATE: 2002-02-06
; PRIOR APPLICATION NUMBER: US 60/268,347
; PRIOR FILING DATE: 2001-02-12
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 467
; TYPE: PRF
; ORGANISM: Bacillus sp.
US-10-068-374-2
```

```
Query Match          10.2%; Score 177; DB 9; Length 467;
Best Local Similarity 25.3%; Pred. No. 1e-07;
Matches 67; Conservative 40; Mismatches 94; Indels 64; Gaps 15;
```

```
QY 29 DMGVVVKDEFDIIKEAGFSHVRIPIRSTHAYAPPYKIMDRFKRVDEVINGALKRGL 88
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 81 DMGIVT-----FRAMYTSSGGYIDP-----SVKEKYKETEVEAIDLGI 120
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 89 AVAINIHHELMNDPEEHKERFLAMKQIADRYKVPETLEFFELINEPHGNLPEKMWNE 148
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 121 YVILDMHILSD--NDPVIYKEAKDFDEMSLXGDPNVI-YEIAEPNG--SDVTWDN 175
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 149 LL-----EALAKVIRSIDKKHTIIGTAEWG---GISALEKLSVPKWKNSIVTTHYNNP 201
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 176 QIKRYAEVIVPIRDNDPNIVIGTWSQDVHHAADNQLADP-----NWYAFHFY--- 228
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 202 EFTHQGAWEWEGSEKWKGRKWSGPDOKHLIEFNFLIEMSKKKRRPIYIGEGAYRKAD 261
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 229 AGTH-GGNLRDQVYAL-----DGAAL-----FVSEWGTG---AATGDGGVF----- 267
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 262 LESRIKWTFSVVREREMKRRMSWAYW 286
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 268 LDEAQVWIDF-----MDERNLSMANW 288
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
```

```
RESULT 12
US-09-863-547B-1
; Sequence 1, Application US/09863547B
; Patent No. US20020128166A1
; GENERAL INFORMATION:
; APPLICANT: Henkel KGaA
; APPLICANT: Hermanns, Lentus B. M.
; APPLICANT: Van Beckhoven, Rudolf F. W. C.
; APPLICANT: Maurer, Karl-Heinz
; APPLICANT: Koltwitz, Beatrix
; APPLICANT: Weiss, Albrecht
; APPLICANT: Van Solingen, Pieter
; TITLE OF INVENTION: Detergents Comprising Cellulases
; FILE REFERENCE: H 1920 A
; CURRENT APPLICATION NUMBER: US/09/863,547B
; CURRENT FILING DATE: 2002-02-13
; PRIOR APPLICATION NUMBER: US 08/945,574
; PRIOR FILING DATE: 1998-02-27
; PRIOR APPLICATION NUMBER: PCT/EP96/01755
; PRIOR FILING DATE: 1996-04-26
; PRIOR APPLICATION NUMBER: US 08/614,115
; PRIOR FILING DATE: 1996-03-12
; PRIOR APPLICATION NUMBER: EP 95201115.3
; PRIOR FILING DATE: 1995-04-28
; NUMBER OF SEQ ID NOS: 2
```


TELEPHONE: 212 867 0123
TELEFAX: 212 867 0298
INFORMATION FOR SEQ ID NO: 13:
SEQUENCE CHARACTERISTICS:
LENGTH: 462 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 13:
US-09-903-185-13

Query Match 9.7%; Score 168; DB 9; Length 462;
Best Local Similarity 24.9%; Pred. No. 6.2e-07;
Matches 66; Conservative 43; Mismatches 92; Indels 64; Gaps 16;

QY 29 DMGVVIRKDEFFDIKEAGFSHVIRPIRMSTHAYAFPPKIMDRFKEVDEVLNGALRGL 88
DB 81 DMGI-----NVFRAMYT-----SSGGYIDDP-----SVKEKVEAVERAIDLDT 120
QY 89 AVAINIHVEELNNDPEHKEERFLAMKQIADRYKDYPTLFEELNEPHNLTPERKWE 148
DB 121 YVLIIDWHIISD--NDPRITYKEEAKDFEDSELYGOVPMYI-YEIANEPNG--SDVTWGN 175
QY 149 LL-----EALKVIRSIDKKHTIIIGTAEWG--GISALEKLSVPKWEKNSIVTHYVNP 201
DB 176 QIKPYAEVPIIRNDPNIIIVGTGWSQDVHHAADNOLADP---NMYAFHFY--- 228
QY 202 EFTHOGAEWEGSEKMGGRKWSGDDOKHLEEFNFIEEWSKKKRPRIYGEFGAYRKAD 261
DB 229 AGTH-GONLRDQVYAL-----DQGAAL---FVSEWGTG---AATGDGVP--- 267
QY 262 LESRIKWTSEVREMEKRRMSWAYW 286
DB 268 LDEAQWIDF---MDERNLSWANW 288

RESULT 15
US-09-927-827-52
Sequence 52, Application US/09927827
Publication No. US20030036176A1
GENERAL INFORMATION:
APPLICANT: Bower, Stanley G.
APPLICANT: Ramsdell, Thomas M.
TITLE OF INVENTION: Directed Genetic Engineering of Xanthomonas campestris
FILE REFERENCE: 38-10(15824)B
CURRENT APPLICATION NUMBER: US/09/927,827
CURRENT FILING DATE: 2001-08-10
PRIOR APPLICATION NUMBER: US 60/279,493
PRIOR FILING DATE: 2001-03-28
NUMBER OF SEQ ID NOS: 69
SEQ ID NO 52
LENGTH: 335
TYPE: PRT
ORGANISM: Xanthomonas campestris
US-09-927-827-52

Query Match 9.2%; Score 159.5; DB 9; Length 335;
Best Local Similarity 21.0%; Pred. No. 2.3e-06;
Matches 69; Conservative 55; Mismatches 133; Indels 71; Gaps 16;

QY 15 GINIGNALEAPNDEMGVVIKDEFFDIKE-----AGFSHVIRPIRMSTHAYAF---FP 64
DB 23 GVNLAGA-EFASSKPGVLNKDYMPASTDYSFAGTGMVIRLPILMERLOPAARGELD 81
QY 65 PYKIMDRFPRVDEVINGALRGLAVAINIHVEEL---MNDPEHKEERFLAMKQIAD 120
DB 82 PAOL-----ALVQAAVARAASGMVTLVDIHNSKYYGYKMGPEVPLATFADLWRRLAV 136
QY 121 RYKDYPTLFEELNEPHNLTPERKWELEALKVIRSIDKKHTIIIGTAEWG----- 175
DB 137 IFNS-DNAVITFGMLNEN-NISASEWAGAQAQIDAIRATGANNLIVPGALMTGAHSMH 194
QY 176 -----ISALEKLSVPKWEKNSIVTHYVNPFEFTHOGAEWEGSEKMGGRKWSGPD 226

DB 195 SLTSDGYSNATALASISDPL--NRFAFEVHOY--LDADSSGTSSVCVSET-IG----- 242
QY 227 DQKHLEEFNFIEEWSKKKRPRIYIGFEGAVRKADLESRIKWTSEVREMEKRR-----W 281
DB 243 -----ADRLRAFTIEWLKTNNKRGFLGEFGTANNVCFNALOG---MLAYMENYADVWLGW 294
QY 282 SW---AYWEEFCGFGVYDTLRKTWNKD 305
DB 295 TWMAGAWMTSYAYNVHP-----NKD 316

Search completed: April 22, 2003, 14:05:39
Job time: 22 secs

GenCore version 5.1.4-p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 26, 2003, 10:49:07 ; Search time 191 Seconds
(without alignments)
11248.201 Million cell updates/sec

Title: US-09-880-729A-1

Perfect score: 954

Sequence: 1 atgggtgtgataccttctga.....gagagatagcatgataa 954

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 2185239 seqs, 112599159 residues.

Word size : 20

Total number of hits satisfying chosen parameters: 23

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database : N.Geneseq_101002.*

```
1: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT:*
2: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT:*
3: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT:*
4: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT:*
5: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT:*
6: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT:*
7: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT:*
8: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT:*
9: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT:*
10: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT:*
11: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT:*
12: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT:*
13: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT:*
14: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:*
15: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT:*
16: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:*
17: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:*
18: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:*
19: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:*
20: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:*
21: /SID52/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:*
22: /SID52/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:*
23: /SID52/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:*
24: /SID52/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	954	100.0	954	20	AAZ21563
2	954	100.0	954	20	AAZ21563
3	954	100.0	954	21	AAZ46759
4	954	100.0	954	22	AAZ46759
5	21	2.2	51	20	AAZ21564
6	21	2.2	51	20	AAZ21564
7	21	2.2	51	21	AAZ46760
8	21	2.2	51	22	AAZ46760
9	21	2.2	530	22	AAZ46760

C 10	20	2.1	849	13	AAO32472	HCV NS2-NS4 clone
C 11	20	2.1	849	13	AAO32483	HCV NS2-NS4 clone
C 12	20	2.1	849	13	AAO32482	HCV NS2-NS4 clone
C 13	20	2.1	849	13	AAO32484	HCV NS2-NS4 clone
C 14	20	2.1	1280	13	AAO32479	HCV NS2-NS4 clone
C 15	20	2.1	1280	13	AAO32494	HCV NS2-NS4 clone
C 16	20	2.1	1280	13	AAO32495	HCV NS2-NS4 clone
C 17	20	2.1	1350	22	AAZ37283	Human musculoskele
C 18	20	2.1	1350	22	AAZ37284	Human musculoskele
C 19	20	2.1	1350	22	AAZ37285	Human musculoskele
C 20	20	2.1	3564	13	AAO32442	HCV NS2-NS4 clone
C 21	20	2.1	3564	13	AAO32501	HCV NS2-NS4 clone
C 22	20	2.1	4235	23	ABL28818	Drosophila melano
C 23	20	2.1	7911	13	AAO32436	HCV antigen clone

ALIGNMENTS

```
RESULT 1
AAZ21563
ID   AAZ21563 standard; cDNA; 954 BP.
AC   AAZ21563;
DT   01-DEC-1999 (first entry)
DE   Carboxymethyl cellulase (CMCase) gene.
XX
XX   Carboxymethyl cellulase; CMCase; thermostable enzyme; hydrolysis;
XX   beta 1,4 glycosidic bond; cellulose; plant biomass conversion; fuel;
XX   chemical; detergent; textile industry; animal feed; waste treatment;
XX   fruit juice extraction; clarification; brewing industry; ss.
XX   Thermotoga maritima.
OS
XX
XX   Key   Location/Qualifiers
XX   CDS   1..954
XX       /tag= a
XX       /product= CMCase
XX       /note= "Carboxymethyl cellulase"
XX
XX   US962258-A.
XX
XX   05-OCT-1999.
XX
XX   23-AUG-1995; 950S-0518615.
XX
XX   23-AUG-1995; 950S-0518615.
XX
XX   (DIVE-) DIVERSA CORP.
XX
XX   Mathur EJ, Lam DE;
XX   WPI: 1999-571269/48.
XX   P-PSDB: AAY39338.
XX
XX   Recombinant polynucleotide sequence encoding a carboxymethyl cellulase
XX   useful for the conversion of plant biomass -
XX
XX   Claim 11; Fig 1; 14pp; English.
XX
XX   This sequence is a carboxymethyl cellulase gene (CMCase). CMCase is a
XX   thermostable enzyme (optimum temperature of 60 + degrees C), that
XX   catalyzes the hydrolysis of the beta 1,4 glycosidic bonds in cellulose.
XX   The enzyme degrades cellulose by the hydrolysis of this bond. The CMCase
XX   enzyme has a molecular weight of approximately 35kd. CMCase is used to
XX   catalyze enzymatic degradation of cellulose and can be used for the
XX   conversion of plant biomass into fuels and chemicals. CMCase can also be
XX   used in detergents, the textile industry, animal feed, waste treatment
XX   and for the clarification and extraction of juices in the fruit juice and
XX   brewing industries. The CMCase gene and 15 bp fragments of it can be used
XX   as probes to isolate nucleic acid molecules encoding CMCase enzymes from
```

CC genomic libraries or to screen for related sequences. Cells expressing
 CC the CMCase polynucleotide can be used as immunogens to produce antibodies
 CC which can then be used to isolate the enzyme from cells which express
 CC CMCase. The antibodies can also be used to screen for similar enzymes
 CC from other organisms and samples.

XX Sequence 954 BP; 335 A; 145 C; 221 G; 253 T; 0 other:

Query Match 100.0%; Score 954; DB 20; Length 954;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 954; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 ATGGGTGTTGATCCTTTTGAAGAACAAATATTGGCAGAGCGCATTAATATAGGAAT 60
Db 1 ATGGGTGTTGATCCTTTTGAAGAACAAATATTGGCAGAGCGCATTAATATAGGAAT 60
QY 61 GCGCTTGAAGACCAATAGAGGAGACTGGGAGTGTGATTAAGATGAGTCTTCGAC 120
Db 61 GCGCTTGAAGACCAATAGAGGAGACTGGGAGTGTGATTAAGATGAGTCTTCGAC 120
QY 121 ATTATAAAGAACCGGTTTCTCATGTTGCAATTCATAGATGAGTACGACGCT 180
Db 121 ATTATAAAGAACCGGTTTCTCATGTTGCAATTCATAGATGAGTACGACGCT 180
QY 181 TACGGCTTCCCTTATTAATCATGATCGCTCTTCAAAAGAGTGAAGTGAATGATA 240
Db 181 TACGGCTTCCCTTATTAATCATGATCGCTCTTCAAAAGAGTGAAGTGAATGATA 240
QY 181 TACGGCTTCCCTTATTAATCATGATCGCTCTTCAAAAGAGTGAAGTGAATGATA 240
Db 181 TACGGCTTCCCTTATTAATCATGATCGCTCTTCAAAAGAGTGAAGTGAATGATA 240
QY 241 AACGAGCCCTGAAAGAGAGACTGGTGTGCTATTAATATTCATCAGAGAGTTA 300
Db 241 AACGAGCCCTGAAAGAGAGACTGGTGTGCTATTAATATTCATCAGAGAGTTA 300
QY 301 ATGAATGATCCAGAAAGACACAAAGAAATTTCTGCTCTTTGGAAACAAATTCGTGAT 360
Db 301 ATGAATGATCCAGAAAGACACAAAGAAATTTCTGCTCTTTGGAAACAAATTCGTGAT 360
QY 361 CCTTTAAAGATCATCCCAAACTATTTTGAATTCGATACATCACTCAGGAAAT 420
Db 361 CCTTTAAAGATCATCCCAAACTATTTTGAATTCGATACATCACTCAGGAAAT 420
QY 421 CTCTACTCCGGAATAATGATGACTGCTTGGAGAACTCTAAAGTTATAGATCAAT 480
Db 421 CTCTACTCCGGAATAATGATGACTGCTTGGAGAACTCTAAAGTTATAGATCAAT 480
QY 481 GCAAAAGACACACTATTAATTAATAGCAGACAGTGAATGGGGGGTATATCTGCCCTTGA 540
Db 481 GCAAAAGACACACTATTAATTAATAGCAGACAGTGAATGGGGGGTATATCTGCCCTTGA 540
QY 541 AAACGTCTGTCTCCCAAAATGGGAAAAAATCTATAGTTACAATTCACACTACATCTCT 600
Db 541 AAACGTCTGTCTCCCAAAATGGGAAAAAATCTATAGTTACAATTCACACTACATCTCT 600
QY 601 TTTCGAATTTACCATCAAGAGAGTGTGGTGAAGATCTGAGAAATGCTGGGAGAGA 660
Db 601 TTTCGAATTTACCATCAAGAGAGTGTGGTGAAGATCTGAGAAATGCTGGGAGAGA 660
QY 661 AAGTGGGATCTCCAGATGATCAGAAACATTTAGTAGAAGAAATTCATTTATAGAGAA 720
Db 661 AAGTGGGATCTCCAGATGATCAGAAACATTTAGTAGAAGAAATTCATTTATAGAGAA 720
QY 721 TGGTCAAAAAAGAACAAAGACCAATTTACATAGTGTGAGTTGGTCTCAGAGAAAGCT 780
Db 721 TGGTCAAAAAAGAACAAAGACCAATTTACATAGTGTGAGTTGGTCTCAGAGAAAGCT 780
QY 781 GACCTTGAATCAAGAAATTAATGACCTCTTGTGCTGGCGAAATGGAGAAAGAGAGA 840
Db 781 GACCTTGAATCAAGAAATTAATGACCTCTTGTGCTGGCGAAATGGAGAAAGAGAGA 840
QY 841 TGGAGCTGGGCATACCTGGAATTTTGTTCGAGTTTGGTGTATGATACCTGAGAGAA 900
Db 841 TGGAGCTGGGCATACCTGGAATTTTGTTCGAGTTTGGTGTATGATACCTGAGAGAA 900
QY 901 ACCTGGAATTAAGATCTTTTGAAGCTTTAATAGAGAGATGACATTGGAATA 954
  
```

Db 901 ACCTGGAATTAAGATCTTTTGAAGCTTTAATAGAGAGATGACATTGGAATA 954

RESULT 2

ID: AAX85879 standard; cDNA; 954 BP.

AXX85879;

10-SEP-1999 (first entry)

cDNA encoding a carboxymethyl cellulase enzyme.

Carboxymethyl cellulase enzyme; hydrolysis; cellulose; plant biomass;

Fuel; Chemical; detergent; textile industry; animal feed;

waste treatment; fruit juice; brewing industry; clarification;

juice extraction; ss.

Thermotoga maritima.

US925749-A.

20-JUL-1999.

23-AUG-1995; 95US-0518615.

23-AUG-1995; 95US-0518615.

24-APR-1998; 98US-0066075.

(DIVE-) DIVERSA CORP.

Lam DE, Mathur EJ;

MP1: 1999-429058/36.

P-PSDB; AAY23764.

New carboxymethyl cellulase enzyme is useful in cellulose digestion

Claim 1; Fig 1A-C; 13pp; English.

The present sequence encodes a carboxymethyl cellulase enzyme from

Thermotoga maritima. Oligonucleotide probes derived from the

polynucleotide sequence can be used to screen a library of genomic

DNA for recovery of the polynucleotide, as PCR primers to synthesize

full length polynucleotides or as diagnostic probes useful for

screening for similar enzymes from other organisms and samples.

The carboxymethyl cellulase enzymes can be used to catalyze hydrolysis

of cellulose for the conversion of plant biomass into fuels and

chemicals, for use in detergents, the textile industry, in animal feed,

waste treatment and in the fruit juice/brewing industry for the

clarification and extraction of juices.

Sequence 954 BP; 335 A; 145 C; 221 G; 253 T; 0 other:

```

QY 1 ATGGGTGTTGATCCTTTTGAAGAACAAATATTGGCAGAGCGCATTAATATAGGAAT 60
Db 1 ATGGGTGTTGATCCTTTTGAAGAACAAATATTGGCAGAGCGCATTAATATAGGAAT 60
QY 61 GCGCTTGAAGACCAATAGAGGAGACTGGGAGTGTGATTAAGATGAGTCTTCGAC 120
Db 61 GCGCTTGAAGACCAATAGAGGAGACTGGGAGTGTGATTAAGATGAGTCTTCGAC 120
QY 121 ATTATAAAGAACCGGTTTCTCATGTTGCAATTCATAGATGAGTACGACGCT 180
Db 121 ATTATAAAGAACCGGTTTCTCATGTTGCAATTCATAGATGAGTACGACGCT 180
QY 181 TACGGCTTCCCTTATTAATCATGATCGCTCTTCAAAAGAGTGAAGTGAATGATA 240
  
```


Db	181	TACGAGTTTCCCTTATTAATATCATGATGCGCTCTTCCAAAAGAGTGGATGCAAGTGATA	240
Qy	241	AACGAGACCCCTGAAAAGAGACGCTGCTGTGCTATATTAATTTTCATCTACGAGAGATT	300
Db	241	AACGAGACCCCTGAAAAGAGACGCTGCTGTGCTATTAATTTTCATCTACGAGAGATT	300
Qy	301	ATGATGATGTCAGAGAAGAACACAAAGAAAGATTCTTGCTCTTTGGAAACAATTTGCTGAT	360
Db	301	ATGATGATGTCAGAGAAGAACACAAAGAAAGATTCTTGCTCTTTGGAAACAATTTGCTGAT	360
Qy	361	CGTTATTAAGACTATVCCCGAAACTATATTTTTGAAATTTCTGAATGAACCTCAGGAAAT	420
Db	361	CGTTATTAAGACTATVCCCGAAACTCTATTTTTTGAATTTCTGAATGAACCTCAGGAAAT	420
Qy	421	CTTACTCCGGAAAAATGGAATGAACGCTCTGAGAAAGCTCTAAAGTTTATGATCAAT	480
Db	421	CTTACTCCGGAAAAATGGAATGAACGCTCTGAGAAAGCTCTAAAGTTTATGATCAAT	480
Qy	481	GACAAAAAGCACACTATATATTAATAGGCACAGCTGAATGGGGGGGTATATCTGCCCTTGA	540
Db	481	GACAAAAAGCACACTATATATTAATAGGCACAGCTGAATGGGGGGGTATATCTGCCCTTGA	540
Qy	541	AAACTGTCTGTCCCAAAATGGGAAAAATTTCTATATGTTACAAATTCACACTACAAATCCT	600
Db	541	AAACTGTCTGTCCCAAAATGGGAAAAATTTCTATAGTTACAAATTCACACTACAAATCCT	600
Qy	601	TTTCAATTTTACCCTCAAGAGAGCTGATGGGTGGAGAGATCTGAGAAATGTTGGGAGAA	660
Db	601	TTTCAATTTTACCCTCAAGAGAGCTGATGGGTGGAGAGATCTGAGAAATGTTGGGAGAA	660
Qy	661	AAGTGGGGATCTCCAGATGATCAGAAAACATTTGATAGAAAGATTCATTTTATAGAGAA	720
Db	661	AAGTGGGGATCTCCAGATGATCAGAAAACATTTGATAGAAAGATTCATTTTATAGAGAA	720
Qy	721	TGGTCAAAAAAGAACAAAGAACCAATTTAATAGGAGATTTGGTGGCTACAGAAAGCT	780
Db	721	TGGTCAAAAAAGAACAAAGAACCAATTTAATAGGAGATTTGGTGGCTACAGAAAGCT	780
Qy	781	GACCTTGAATCAAGAAATTAATGAGACCCTCTTTGTGCTTCCGGAATGGAGAAAGAGAGA	840
Db	781	GACCTTGAATCAAGAAATTAATGAGACCCTCTTTGTGCTTCCGGAATGGAGAAAGAGAGA	840
Qy	841	TGGAGCTGGGGATCTCTGGAATTTTGTCCGGTTTTGGCTTTTATGATATCTCTGAGAAAA	900
Db	841	TGGAGCTGGGGATCTCTGGAATTTTGTCCGGTTTTGGCTTTTATGATATCTCTGAGAAAA	900
Qy	901	ACCTGGAATTAAGATCTTTTAGAAGCTTTAATAGAGAGATAGCATTAATAA	954
Db	901	ACCTGGAATTAAGATCTTTTAGAAGCTTTAATAGAGAGATAGCATTAATAA	954
RESULT 3			
AAZ46759	AAZ46759 standard; cDNA; 954 BP.		
AAZ46759:			
31-MAR-2000	(first entry)		
T. maritima	CMCase enzyme encoding cDNA.		
Endoglucanase; carboxymethyl cellulase; cellulose; CMCase; plant biomass;			
beta-1,4-glycosidic bond; detergent; textile; plant gum; animal feed;			
waste product; fuel; brewing; thermostable; ss.			
Thermotoga maritima.			
Location/Qualifiers			
1.954			
/+tag- a			
/product= "CMCase"			
US6008032-A			

XX	28-DEC-1999.
PD	
XX	
PF	16-OCT-1997; 97US-0951889.
XX	
PR	23-AUG-1995; 95US-0518615.
XX	
PA	(DIVE-) DIVERSA CORP.
XX	
PI	Mathur EJ, Lam DE;
XX	
WP	WPI; 2000-105335/09.
DR	P-PSDB; AAY56814.
XX	
PT	Endoglucanase with carboxymethyl cellulase activity from <i>Thermotoga</i>
PT	<i>maritima</i> , useful for degrading biomass to fuels and chemicals
XX	
PS	Disclosure; Fig 1A-C; 14pp; English.
XX	
CC	This cDNA encodes a T. <i>maritima</i> endoglucanase with carboxymethyl
CC	cellulase activity. The enzyme CMCase is thermostable and used to degrade
CC	carboxymethyl cellulose (CMC) or more generally any oligosaccharide that
CC	contains beta-1,4-glycosidic bonds, particularly in detergents, textiles,
CC	gum arabic, animal feed, plant biomass or waste products. The CMCase enzyme
CC	is especially used to convert plant biomass to chemicals and fuels, and
CC	also for clarification and extraction of juice (e.g. in the brewing
CC	industry). CMCcase can also be used to raise specific antibodies.
XX	
50	Sequence 954 BP; 335 A; 145 C; 221 G; 253 T; 0 other;

Query Match	Similarity	100.0%	Score 954:	DB 21:	Length 954:
Best Local	Similarity	100.0%	Pred. NO. 0:		
Matches	954:	Conservative	0:	Mismatches	0:
				Indels	Gaps
QY	1	ATGGCTGTTGATCCCTTTTGAAGAAGCAAAATATTTGGAGAGAGCATTTATATAGAAAT	60		
Db	1	ATGGCTGTTGATCCCTTTTGAAGAAGCAAAATATTTGGAGAGAGCATTTATATAGAAAT	60		
QY	61	GGCGTTGAAGCACCAATATGAGGAGACTGGGAGTGGATATTAAGAATGATTTCTTGCAC	120		
Db	61	GGCGTTGAAGCACCAATATGAGGAGACTGGGAGTGGATATTAAGAATGATTTCTTGCAC	120		
QY	121	ATTATTAAGAAGAGCGGTTTCTCATGTGTGCAATTTCCATAAGATGAGTACGCGACGT	180		
Db	121	ATTATTAAGAAGAGCGGTTTCTCATGTGTGCAATTTCCATAAGATGAGTACGCGACGT	180		
QY	181	TACGGCTTTCCTCCTTATTAATAATCATGATGATCGCTTCTTCAAAAAGATGGATGAAGTGATA	240		
Db	181	TACGGCTTTCCTCCTTATTAATAATCATGATGATCGCTTCTTCAAAAAGATGGATGAAGTGATA	240		
QY	241	AACGAGAGCCCTGAAAAGAGAGACTGGCTGTGCTATTAATATTTCAATCAGTACGAGAGTTA	300		
Db	241	AACGAGAGCCCTGAAAAGAGAGACTGGCTGTGCTATTAATTAATTTCAATCAGTACGAGAGTTA	300		
QY	301	ATGATTTGATTCGAGAGACACACAGAAGAAATTTCTTGCTCTTTGGAAACCAATTTGCTGAT	360		
Db	301	ATGATTTGATTCGAGAGACACACAGAAGAAATTTCTTGCTCTTTGGAAACCAATTTGCTGAT	360		
QY	361	CGTTATTAAGACTATCCGGAACCTCTATTTTTTGAATTTGGAATTCGAATGAACCTCACGGAAAT	420		
Db	361	CGTTATTAAGACTATCCGGAACCTCTATTTTTTGAATTTGGAATTCGAATGAACCTCACGGAAAT	420		
QY	421	CTTTACTCCGGAAAAATGTAATGTAAGTACCTGTGAGGAAGCTCTTAAAGTTATTAAGATCAAT	480		
Db	421	CTTTACTCCGGAAAAATGTAATGTAAGTACCTGTGAGGAAGCTCTTAAAGTTATTAAGATCAAT	480		
QY	481	GACAAAGAAGCACATATATATATAGCAGCAGCTGAATGGGGGGATATTTCTGCCCTTGA	540		
Db	481	GACAAAGAAGCACATATATATATAGCAGCAGCTGAATGGGGGGATATTTCTGCCCTTGA	540		
QY	541	AAACTGTCTGCCAAATGGGAAAAAAATTTCTATAGTTACATTTCACTACTACAATCCT	600		
Db	541	AAACTGTCTGCCAAATGGGAAAAAAATTTCTATAGTTACATTTCACTACTACAATCCT	600		

QY	601	TTCCAAATTTTACCCATTC	CAAGGAGCTGAGTGGGTGCAAGCATCTGAAATGCTTGGGAGAG	660
Db	601	TTCCAAATTTTACCCATTC <th>CAAGGAGCTGAGTGGGTGCAAGCATCTGAAATGCTTGGGAGAG</th> <th>660</th>	CAAGGAGCTGAGTGGGTGCAAGCATCTGAAATGCTTGGGAGAG	660
QY	661	AAGTGGGATCTCCAGATGAT	TCGAAACATTTTGATAGCAATTTCAATTTTATAGAAGAA	720
Db	661	AAGTGGGATCTCCAGATGAT	TCGAAACATTTTGATAGCAATTTTCAATTTTATAGAAGAA	720
QY	721	TGTCACAAAAAAGAACAAAG	ACCAATTTTACATAGTAGAGTTTGGTGCCTACAGAAAAGCT	780
Db	721	TGTCACAAAAAAGAACAAAG	ACCAATTTTACATAGTAGAGTTTGGTGCCTACAGAAAAGCT	780
QY	781	GACCTTGAAATCAAGAAATTA	ATGAGACCTCTTGCGTCGCCAAATGGAGAAAAGGAGA	840
Db	781	GACCTTGAAATCAAGAAATTA	ATGAGACCTCTTGCGTCGCCAAATGGAGAAAAGGAGA	840
QY	841	TGAGACTGGGATCTACTGGGA	ATTTTGTTCGCGTTTGGTGTGTTATGATCTCGAGAAA	900
Db	841	TGAGACTGGGATCTACTGGGA	ATTTTGTTCGCGTTTGGTGTGTTATGATCTCGAGAAA	900
QY	901	ACCTGGAATTAAGATCTTTTGA	AGCCTTTAATAGAGAGATAGCATGATTA	954
Db	901	ACCTGGAATTAAGATCTTTTGA	AGCCTTTAATAGAGAGATAGCATGATTA	954
RESULT 4				
	AD15243	ID AD15243 standard; DNA; 954 BP.		
XX	AC	AAD15243;		
XX	DT	15-NOV-2001 (first entry)		
XX	DE	Thermotoga maritima CMCase enzyme encoding DNA.		
XX	KW	Carboxymethyl cellulose; CMCase; plant biomass; animal feed; detergent;		
XX	KW	textile industry; waste treatment; brewing industry; immunogen;		
XX	KW	fruit juice industry; ds.		
XX	OS	Thermotoga maritima.		
XX	EH	Key Location/Qualifiers		
XX	FT	CDS 1..954		
XX	FT	/tag= "T. maritima CMCase enzyme "		
XX	PN	US6245547-B1.		
XX	PD	12-JUN-2001.		
XX	PE	27-DEC-1999; 99US-0472857.		
XX	PR	23-AUG-1995; 95US-0518615.		
XX	PR	16-OCT-1997; 97US-0951889.		
XX	PA	(DIVE-) DIVERSA CORP.		
XX	PI	Mathur EJ, Lam DE;		
XX	DR	WPI; 2001-528380/58.		
XX	DR	P-PDB; AAE08543.		
XX	PT	Polynucleotide isolated from thermophile eubacteria useful for		
XX	PT	degradation of cellulose encodes an enzyme having a fully defined amino		
XX	PT	acid sequence		
XX	PS	Claim 3; Fig 1; 14pp; English.		
XX	CC	The present sequence is a DNA coding for a thermostable enzyme, an		
XX	CC	endoglucanase having carboxymethyl cellulose activity referred to as		
XX	CC	(CMCase). The CMCase enzyme is useful for the degradation of cellulose,		
XX	CC	for conversion of plant biomass into fuels and chemicals, in the		

Query Match	100.0%	Score 954:	DB 22:	Length 954:
Best Local Similarity	100.0%	Pred. No. 0:		
Matches 954:	Conservative 0:	Mismatches 0:	Indels 0:	Gaps 0:
Oy	1	ATGGGCTGTTCATCCCTTTTGAAGGAACAATAATTGGGAAGAGCCATTATATAGCAAT	60	
Db	1	ATGGGCTGTTCATCCCTTTTGAAGGAACAATAATTGGGAAGAGCCATTATATAGCAAT	60	
Oy	61	GGCGTTGAAGACCACAATGAGGAGACTGGGGAGTGGTATAAAGATGAGCTTTTGAC	120	
Db	61	GGCGTTGAAGACCACAATGAGGAGACTGGGGAGTGGTATAAAGATGAGCTTTTGAC	120	
Oy	121	ATTATTAAGAAAGACGGCTTCTCATGTTTGCATTCACATTAAGATGAGTACCGACCT	180	
Db	121	ATTATTAAGAAAGACGGCTTCTCATGTTTGCATTCACATTAAGATGAGTACCGACCT	180	
Oy	181	TACGGCTTCTCTTATTAATATATGATACGCTTCTTAAAGAGTGGATGAAGTGATA	240	
Db	181	TACGGCTTCTCTTATTAATATATGATACGCTTCTTAAAGAGTGGATGAAGTGATA	240	
Oy	241	AACGGAGCCCTGAANAAGAGAGCTGGCTGTTCATTAATATTCATCCTACGAGAGTTA	300	
Db	241	AACGGAGCCCTGAANAAGAGAGCTGGCTGTTCATTAATATTCATCCTACGAGAGTTA	300	
Oy	301	ATGATATGATCCAGAGAACACACAGAGAAAGATTCTTGCTCTTTGGAAACCAATTGCTGAT	360	
Db	301	ATGATATGATCCAGAGAACACACAGAGAAAGATTCTTGCTCTTTGGAAACCAATTGCTGAT	360	
Oy	361	CGTTATTAAGACTATCCCGGAACCTCTATTTTTGAATTTCTGATGTAACCTCCAGCAAT	420	
Db	361	CGTTATTAAGACTATCCCGGAACCTCTATTTTTGAATTTCTGATGTAACCTCCAGCAAT	420	
Oy	421	CTTACTCCGGAAAAATGGAATGAACTGCTTGAGAGACTTAAAGCTTTAATGATCAAT	480	
Db	421	CTTACTCCGGAAAAATGGAATGAACTGCTTGAGAGACTTAAAGCTTTAATGATCAAT	480	
Oy	481	GACAAAACACACATATATTAATAGGACACAGCTGAATGGGGGGTATATCTGCCCTTGA	540	
Db	481	GACAAAACACACATATATTAATAGGACACAGCTGAATGGGGGGTATATCTGCCCTTGA	540	
Oy	541	AAACTGTCTGCCAAAAATGGGAAAAAAATCTTATGTTCAATTCATCTACTACTCAATCT	600	
Db	541	AAACTGTCTGCCAAAAATGGGAAAAAAATCTTATGTTCAATTCATCTACTACTCAATCT	600	
Oy	601	TTTGCATTTTACCCATCAAGAGAGCTGAGGGGTGGAGAGCTCAGAAATGGTGGGAGA	660	
Db	601	TTTGCATTTTACCCATCAAGAGAGCTGAGGGGTGGAGAGCTCAGAAATGGTGGGAGA	660	
Oy	661	AAGTGGGATCTCCAGATGATCAGAAACATTTGATAGAGATTTCAATTTATAGAGAA	720	
Db	661	AAGTGGGATCTCCAGATGATCAGAAACATTTGATAGAGATTTCAATTTATAGAGAA	720	
Oy	721	TGTCACAAAAAGAACAAAGACCAATTTACATAGGTGGTGGTGGCTCCACAAAGCT	780	
Db	721	TGTCACAAAAAGAACAAAGACCAATTTACATAGGTGGTGGTGGCTCCACAAAGCT	780	
Oy	781	GACCTTAATCAGATTAATAATGAGACTCTTGTGCTTGGCGCAATGAGAAAGAGAA	840	
Db	781	GACCTTAATCAGATTAATAATGAGACTCTTGTGCTTGGCGCAATGAGAAAGAGAA	840	
Oy	841	TGAGACTGGGCTACTGGGAAATTTGGTCCGTTTGGTGTATTGATACCTGAGAGAA	900	
Db	841	TGAGACTGGGCTACTGGGAAATTTGGTCCGTTTGGTGTATTGATACCTGAGAGAA	900	
Oy	901	ACCTGGAATTAAGATCTTTTAGAGCTTAAATAGAGAGAGATACATTTGAATAA	954	

Db 901 ACCTGGAATAAGATCTTTAGAACGTTTAAATAGAGAGATGACATTGATATA 954

RESULT 5

AA221564
ID AA221564 standard; cDNA, 51 BP.

XX AA221564;

XX 01-DEC-1999 (first entry)

XX PCR primer for the amplification of carboxymethyl cellulase gene.

XX Carboxymethyl cellulase; CMCase; thermostable enzyme; hydrolysis;

KW beta 1,4 glycosidic bond; cellulose; plant biomass conversion; fuel;

KW chemical; detergent; textile industry; animal feed; waste treatment;

KW fruit juice extraction; clarification; brewing industry; PCR primer; ss.

XX Synthetic.

OS Thermotoga maritima.

PN US5962258-A.

XX 05-OCT-1999.

XX 23-AUG-1995; 95US-0518615.

XX 23-AUG-1995; 95US-0518615.

XX (DIVE-) DIVERSA CORP.

PA Mathur EJ, Lam DE;

XX WPI; 1999-571269/48.

XX Recombinant polynucleotide sequence encoding a carboxymethyl cellulase

XX useful for the conversion of plant biomass -

XX Example 1; Column 14; 14pp; English.

XX PCR primers AA221564-221565 are used to amplify the carboxymethyl.

XX cellulase gene (CMCase). CMCase is a thermostable enzyme (optimum

XX temperature of 60 + degrees C), that catalyses the hydrolysis of the beta

XX 1,4 glycosidic bonds in cellulose. The enzyme degrades cellulose by the

XX hydrolysis of this bond. The CMCase enzyme has a molecular weight of

XX approximately 35kD. CMCase is used to catalyse enzymatic degradation of

XX cellulose and can be used for the conversion of plant biomass into fuels

XX and chemicals. CMCase can also be used in detergents, the textile

XX industry, animal feed, waste treatment and for the clarification and

XX extraction of juices in the fruit juice and brewing industries. The

XX CMCase gene and 15 bp fragments of it can be used as probes to isolate

XX nucleic acid molecules encoding CMCase enzymes from genomic libraries or

XX to screen for related sequences. Cells expressing the CMCase

XX polynucleotide can be used as immunogens to produce antibodies which can

XX then be used to isolate the enzyme from cells which express CMCase. The

XX antibodies can also be used to screen for similar enzymes from other

XX organisms and samples.

XX Sequence 51 BP; 14 A; 6 C; 14 G; 17 T; 0 other;

XX Query Match 2.2%; Score 21; DB 20; Length 51;

XX Best Local Similarity 100.0%; Pred. No. 3.4;

XX Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

XX QY 1 ATGGGTGTGATCCTTTGAA 21

XX ||||||||||||||||||

XX DB 31 ATGGGTGTGATCCTTTGAA 51

XX RESULT 6

XX AA221564

XX ID AA221564 standard; cDNA; 51 BP.

XX AAX85880;

XX 10-SEP-1999 (first entry)

XX PCR primer used to amplify carboxymethyl cellulase enzyme cDNA.

XX Carboxymethyl cellulase enzyme; hydrolysis; cellulose; plant biomass;

KW fuel; chemical; detergent; textile industry; animal feed;

KW waste treatment; fruit juice; brewing industry; clarification;

KW juice extraction; PCR primer; ss.

XX Thermotoga maritima.

OS US5925749-A.

PN 20-JUL-1999.

XX 23-AUG-1995; 95US-0518615.

XX 23-AUG-1995; 95US-0518615.

XX (DIVE-) DIVERSA CORP.

PA Lam DE, Mathur EJ;

XX WPI; 1999-429058/36.

XX New carboxymethyl cellulase enzyme is useful in cellulose digestion

XX Example 1; Column 14; 13pp; English.

XX PCR primers AAX85880-81 were used to amplify cDNA encoding a

XX carboxymethyl cellulase enzyme from Thermotoga maritima. Oligonucleotide

XX probes derived from the polynucleotide sequence can be used to screen a

XX library of genomic DNA for recovery of the polynucleotide, as PCR primers

XX CC for synthesise full length polynucleotides or as diagnostic probes useful

XX CC for screening for similar enzymes from other organisms and samples.

XX CC The carboxymethyl cellulase enzymes can be used to catalyse hydrolysis

XX CC of cellulose for the conversion of plant biomass into fuels and

XX CC chemicals, for use in detergents, the textile industry, in animal feed,

XX CC waste treatment and in the fruit juice/brewing industry for the

XX CC clarification and extraction of juices.

XX Sequence 51 BP; 14 A; 6 C; 14 G; 17 T; 0 other;

XX Query Match 2.2%; Score 21; DB 20; Length 51;

XX Best Local Similarity 100.0%; Pred. No. 3.4;

XX Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

XX QY 1 ATGGGTGTGATCCTTTGAA 21

XX ||||||||||||||||||

XX DB 31 ATGGGTGTGATCCTTTGAA 51

XX RESULT 7

XX AA246760

XX ID AA246760 standard; DNA; 51 BP.

XX AC AA246760;

XX DT 31-MAR-2000 (first entry)

XX T. maritima CMCase cDNA amplifying primer.

XX Endoglucanase; carboxymethyl cellulase; cellulose; CMCase; plant biomass;

KW beta-1,4-glycosidic bond; detergent; textile; guar gum; animal feed;

KW waste product; fuel; brewing; thermostable; PCR primer; ss.

XX Thermotoga maritima.

OS US6008032-A.

PN

```

XX 28-DEC-1999.
PD 16-OCT-1997; 97US-0951889.
XX 23-AUG-1995; 95US-0518615.
XX (DIVE-) DIVERSA CORP.
PA
XX Mathur EJ, Lam DE;
XX WPI: 2000-105335/09.
XX
XX Endoglucanase with carboxymethyl cellulase activity from Thermotoga
PT maritima, useful for degrading biomass to fuels and chemicals
XX
XX Disclosure: Col 17; 14pp; English.
XX
XX Sequences AA246760-761 represents PCR primers for the amplification of
CC the cDNA encoding a T. maritima endoglucanase with carboxymethyl
CC cellulase activity. The enzyme CMCase is thermostable and used to
CC degrade carboxymethyl cellulose (CMC) or more generally any
CC oligosaccharide that contains beta-1,4-glycosidic bonds, particularly in
CC detergents, textiles, guar gum, animal feed, plant biomass or waste
CC products. The CMCase enzyme is especially used to convert plant biomass
CC to chemicals and fuels, and also for clarification and extraction of
CC juice (e.g. in the brewing industry). CMCase can also be used to raise
CC specific antibodies.
XX
XX Sequence 51 BP; 14 A; 6 C; 14 G; 17 T; 0 other;
SQ
Query Match 2.2%; Score 21; DB 21; Length 51;
Best Local Similarity 100.0%; Pred. No. 3.4;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 ATGGGTGTGATCCTTTGAA 21
DB 31 ATGGGTGTGATCCTTTGAA 51
RESULT 8
AAD15244
ID AAD15244 standard; DNA; 51 BP.
XX
XX AAD15244;
AC
XX 15-NOV-2001 (first entry)
DT
XX
XX Thermotoga maritima CMCase DNA amplifying primer #1.
DE
XX
XX Carboxymethyl cellulase; CMCase; plant biomass; animal feed; detergent;
KW textile industry; waste treatment; brewing industry; immunogen;
KW fruit juice industry; PCR primer; ss.
XX
XX Thermotoga maritima.
OS
XX
XX Key Location/Qualifiers
FH misc_feature 28
FT /+tag= a
FT /note= "This base is absent in the sequence shown
FT as SEQ ID NO:3 in Column 14 of the specification"
XX
XX US6245547-B1.
PN
XX
XX 12-JUN-2001.
PD
XX
XX 27-DEC-1999; 99US-0472857.
XX
XX 23-AUG-1995; 95US-0518615.
XX
XX 16-OCT-1997; 97US-0951889.
XX
XX (DIVE-) DIVERSA CORP.
PA
XX

```

```

PI Mathur EJ, Lam DE;
XX
XX WPI: 2001-528380/58.
XX
XX Polynucleotide isolated from thermophile eubacteria useful for
PT degradation of cellulose encodes an enzyme having a fully defined amino
PT acid sequence
XX
XX Example 1; Column 17-18; 14pp; English.
XX
XX The present sequence is a PCR primer used to amplify a DNA coding for
XX a thermostable enzyme, an endoglucanase having carboxymethyl cellulase
XX activity referred to as (CMCase).
XX
XX The CMCase enzyme is useful for the degradation of cellulose,
XX for conversion of plant biomass into fuels and chemicals, in the
XX textile industry, in detergents, in animal feed, in waste treatment,
XX in fruit juice/brewing industry for clarification and extraction of
XX juices, to generate probes for identifying sequences encoding similar
XX enzymes from other organisms, as immunogen to produce monoclonal or
XX polyclonal antibodies.
XX
XX Sequence 51 BP; 14 A; 6 C; 14 G; 17 T; 0 other;
SQ
Query Match 2.2%; Score 21; DB 22; Length 51;
Best Local Similarity 100.0%; Pred. No. 3.4;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 ATGGGTGTGATCCTTTGAA 21
DB 31 ATGGGTGTGATCCTTTGAA 51
RESULT 9
AAL35482
ID AAL35482 standard; CDNA; 530 BP.
XX
XX AAL35482;
AC
XX 08-JAN-2002 (first entry)
DT
XX
XX Human musculoskeletal system related polynucleotide SEQ ID NO 824.
DE
XX
XX Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;
KW antiallergic; hepatotropic; antidiabetic; antiinflammatory; anticancer;
KW vulnerrary; anticonvulsant; antibacterial; antifungal; antiparasitic;
KW cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;
KW neurological disease; infection; human; secreted protein;
KW musculoskeletal system; ss.
XX
XX Homo sapiens.
OS
XX
XX WO200155367-A1.
XX
XX 02-AUG-2001;
PD
XX
XX 17-JAN-2001; 2001WO-US01338.
XX
XX 31-JAN-2000; 2000US-0179065.
XX
XX 04-FEB-2000; 2000US-0180628.
XX
XX 24-FEB-2000; 2000US-0184664.
XX
XX 02-MAR-2000; 2000US-0186350.
XX
XX 16-MAR-2000; 2000US-0189874.
XX
XX 17-MAR-2000; 2000US-0190076.
XX
XX 18-APR-2000; 2000US-0198123.
XX
XX 19-MAY-2000; 2000US-0205153.
XX
XX 07-JUN-2000; 2000US-0209467.
XX
XX 28-JUN-2000; 2000US-0214886.
XX
XX 30-JUN-2000; 2000US-0215135.
XX
XX 07-JUL-2000; 2000US-0216647.
XX
XX 07-JUL-2000; 2000US-0216880.
XX
XX 11-JUL-2000; 2000US-0217487.
XX
XX 11-JUL-2000; 2000US-0217496.
XX
XX 14-JUL-2000; 2000US-0218290.
XX

```

PR 26-JUL-2000; 2000US-0220963.
 PR 26-JUL-2000; 2000US-0220964.
 PR 14-AUG-2000; 2000US-0224518.
 PR 14-AUG-2000; 2000US-0224519.
 PR 14-AUG-2000; 2000US-0225213.
 PR 14-AUG-2000; 2000US-0225214.
 PR 14-AUG-2000; 2000US-0225266.
 PR 14-AUG-2000; 2000US-0225267.
 PR 14-AUG-2000; 2000US-0225268.
 PR 14-AUG-2000; 2000US-0225270.
 PR 14-AUG-2000; 2000US-0225477.
 PR 14-AUG-2000; 2000US-0225757.
 PR 14-AUG-2000; 2000US-0225758.
 PR 14-AUG-2000; 2000US-0225759.
 PR 18-AUG-2000; 2000US-0226279.
 PR 22-AUG-2000; 2000US-0226681.
 PR 22-AUG-2000; 2000US-0226686.
 PR 22-AUG-2000; 2000US-0227182.
 PR 23-AUG-2000; 2000US-0227009.
 PR 30-AUG-2000; 2000US-0228924.
 PR 01-SEP-2000; 2000US-0229287.
 PR 01-SEP-2000; 2000US-0229343.
 PR 01-SEP-2000; 2000US-0229344.
 PR 01-SEP-2000; 2000US-0229345.
 PR 05-SEP-2000; 2000US-0229509.
 PR 05-SEP-2000; 2000US-0229513.
 PR 06-SEP-2000; 2000US-0230437.
 PR 06-SEP-2000; 2000US-0230438.
 PR 08-SEP-2000; 2000US-0231242.
 PR 08-SEP-2000; 2000US-0231243.
 PR 08-SEP-2000; 2000US-0231244.
 PR 08-SEP-2000; 2000US-0231413.
 PR 08-SEP-2000; 2000US-0231414.
 PR 08-SEP-2000; 2000US-0232080.
 PR 08-SEP-2000; 2000US-0232081.
 PR 12-SEP-2000; 2000US-0231968.
 PR 14-SEP-2000; 2000US-0232397.
 PR 14-SEP-2000; 2000US-0232398.
 PR 14-SEP-2000; 2000US-0232399.
 PR 14-SEP-2000; 2000US-0232400.
 PR 14-SEP-2000; 2000US-0232401.
 PR 14-SEP-2000; 2000US-0233063.
 PR 14-SEP-2000; 2000US-0233064.
 PR 14-SEP-2000; 2000US-0233065.
 PR 21-SEP-2000; 2000US-0234223.
 PR 21-SEP-2000; 2000US-0234274.
 PR 25-SEP-2000; 2000US-0234997.
 PR 25-SEP-2000; 2000US-0234998.
 PR 26-SEP-2000; 2000US-0235484.
 PR 27-SEP-2000; 2000US-0235834.
 PR 27-SEP-2000; 2000US-0235836.
 PR 29-SEP-2000; 2000US-0236327.
 PR 29-SEP-2000; 2000US-0236367.
 PR 29-SEP-2000; 2000US-0236368.
 PR 29-SEP-2000; 2000US-0236369.
 PR 29-SEP-2000; 2000US-0236370.
 PR 02-OCT-2000; 2000US-0236802.
 PR 02-OCT-2000; 2000US-0237037.
 PR 02-OCT-2000; 2000US-0237038.
 PR 02-OCT-2000; 2000US-0237039.
 PR 02-OCT-2000; 2000US-0237040.
 PR 13-OCT-2000; 2000US-0239937.
 PR 20-OCT-2000; 2000US-0240960.
 PR 20-OCT-2000; 2000US-0241221.
 PR 20-OCT-2000; 2000US-0241785.
 PR 20-OCT-2000; 2000US-0241786.
 PR 20-OCT-2000; 2000US-0241787.
 PR 20-OCT-2000; 2000US-0241808.
 PR 20-OCT-2000; 2000US-0241809.
 PR 20-OCT-2000; 2000US-0241826.
 PR 01-NOV-2000; 2000US-0244617.
 PR 08-NOV-2000; 2000US-0244674.

PR 08-NOV-2000; 2000US-0246475.
 PR 08-NOV-2000; 2000US-0246476.
 PR 08-NOV-2000; 2000US-0246477.
 PR 08-NOV-2000; 2000US-0246478.
 PR 08-NOV-2000; 2000US-0246523.
 PR 08-NOV-2000; 2000US-0246524.
 PR 08-NOV-2000; 2000US-0246525.
 PR 08-NOV-2000; 2000US-0246526.
 PR 08-NOV-2000; 2000US-0246527.
 PR 08-NOV-2000; 2000US-0246528.
 PR 08-NOV-2000; 2000US-0246532.
 PR 08-NOV-2000; 2000US-0246609.
 PR 08-NOV-2000; 2000US-0246610.
 PR 08-NOV-2000; 2000US-0246611.
 PR 08-NOV-2000; 2000US-0246613.
 PR 17-NOV-2000; 2000US-0249207.
 PR 17-NOV-2000; 2000US-0249208.
 PR 17-NOV-2000; 2000US-0249209.
 PR 17-NOV-2000; 2000US-0249210.
 PR 17-NOV-2000; 2000US-0249211.
 PR 17-NOV-2000; 2000US-0249212.
 PR 17-NOV-2000; 2000US-0249213.
 PR 17-NOV-2000; 2000US-0249214.
 PR 17-NOV-2000; 2000US-0249215.
 PR 17-NOV-2000; 2000US-0249216.
 PR 17-NOV-2000; 2000US-0249217.
 PR 17-NOV-2000; 2000US-0249218.
 PR 17-NOV-2000; 2000US-0249244.
 PR 17-NOV-2000; 2000US-0249245.
 PR 17-NOV-2000; 2000US-0249264.
 PR 17-NOV-2000; 2000US-0249265.
 PR 17-NOV-2000; 2000US-0249297.
 PR 17-NOV-2000; 2000US-0249299.
 PR 17-NOV-2000; 2000US-0249300.
 PR 01-DEC-2000; 2000US-0250160.
 PR 01-DEC-2000; 2000US-0250391.
 PR 05-DEC-2000; 2000US-0251030.
 PR 05-DEC-2000; 2000US-0251988.
 PR 05-DEC-2000; 2000US-0256719.
 PR 06-DEC-2000; 2000US-0251479.
 PR 08-DEC-2000; 2000US-0251856.
 PR 08-DEC-2000; 2000US-0251868.
 PR 08-DEC-2000; 2000US-0251869.
 PR 08-DEC-2000; 2000US-0251899.
 PR 08-DEC-2000; 2000US-0251990.
 PR 11-DEC-2000; 2000US-0254097.
 PR 05-JAN-2001; 2001US-0259678.
 PR (HUMA-) HUMAN GENOME SCI INC.
 PR PA
 PR XX
 PR XX
 PR PI
 PR Rosen CA, Barash SC, Ruben SM;
 PR WPI, 2001-451937/48.
 PR P-PSDB; ABB03900.
 PR XX
 PT Isolated polypeptide for treating, preventing and/ or prognosing
 PT disorders related to the musculoskeletal system including
 PT musculoskeletal cancers and also for testing and detection e.g.
 PT diagnosis -
 PT PT
 XX
 XX
 XX
 XX
 CC Claim 1; SEQ ID NO 824; 781pp + Sequence Listing; English.
 CC The invention relates to novel genes (AAU34669-AAU37666) and proteins
 CC (ABB03087-ABB04109) associated with the musculoskeletal system useful
 CC for preventing, treating or ameliorating medical conditions e.g. by
 CC protein or gene therapy. The genes are isolated from a range of human
 CC tissues disclosed in the specification. The nucleic acids, proteins,
 CC antibodies and (anti)agonists are useful in the diagnosis, treatment
 CC and prevention of: (a) cancer, e.g. breast and ovarian cancer and
 CC other cancers of the adrenal gland, bone, bone marrow, breast,
 CC gastrointestinal tract, liver, lung, or urogenital; (b) immune
 CC disorders e.g. Addison's disease, allergies, autoimmune haemolytic
 CC anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's disease,


```
RESULT 12
AA032482/c
ID AA032482 standard; DNA: 849 BP.
XX
XX
AC AA032482;
XX
XX
DT 26-APR-1993 (first entry)
XX
DE HCV NS2-NS4 clone MX25-1.
XX
XX
KM Clone; polypeptide; NS2-NS4; Hepatitis C; Virus; HCV; serum; HC;
KM transcriptase; cDNA; primer; allele; ss.
XX
XX
OS Hepatitis C virus.
XX
XX
PN EP518313-A.
XX
XX
PD 16-DEC-1992.
XX
XX
PF 11-JUN-1992; 92EP-0109812.
XX
XX
PR 11-JUN-1991; 91JP-0139268.
PR 12-JUL-1991; 91JP-0172794.
PR 07-OCT-1991; 91JP-0287008.
PR 16-DEC-1991; 91JP-0332329.
PR 20-APR-1992; 92JP-0099957.
XX
XX
PA (MITU ) MITSUBISHI KASEI CORP.
XX
XX
PI Hayashi N, Honda Y, Murakami T, Seki M, Takahashi K;
PI Teranishi Y;
XX
XX
DR WPI; 1992-417213/51.
DR P-PSDB; AAR29851.
XX
XX
PT New hepatitis C virus gene and its encoded protein - used for
PT diagnosing and vaccinating against hepatitis C virus infections
XX
XX
PS Disclosure; Page 146-47; 305pp; English.
XX
XX
CC The sequences given in AA032472-82 and AA032442 are various clones which
CC encode the NS2-NS4 regions of the Hepatitis C Virus (HCV) gene of
CC the invention. These sequences were isolated from the serum of a
CC patient suffering from hepatitis C (HC). The isolated RNA sequences
CC were converted into cDNA using transcriptase in the presence of one
CC of the primer sequences given in AA032553-64. The sequences were
CC then amplified using primer pairs. The cDNA sequences isolated
CC represent different alleles of the same region of the HCV gene.
CC Sequence comparisons of these clones showed that it is possible for a
CC patient to carry more than one HCV strain at one time. See also
CC AA032436.
XX
XX
SQ Sequence 849 BP; 146 A; 245 C; 249 G; 209 T; 0 other;
XX
XX
Query Match 2.1%; Score 20; DB 13; Length 849;
Best Local Similarity 100.0%; Pred. No. 10;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
QY 311 CAGAGAACACAGGAAGA 330
DB 125 CAGAGAACACAGGAAGA 106
XX
XX
RESULT 13
AA032484/c
ID AA032484 standard; DNA: 849 BP.
XX
XX
AC AA032484;
XX
XX
DT 26-APR-1993 (first entry)
XX
XX
DE HCV NS2-NS4 clone MX25-3.
XX
XX
```

```
KM Clone; polypeptide; NS2-NS4; Hepatitis C; Virus; HCV; serum; HC;
KM transcriptase; cDNA; primer; allele; ss.
XX
XX
XX Hepatitis C virus.
XX
XX
XX EP518313-A.
XX
XX
XX 16-DEC-1992.
XX
XX
XX 11-JUN-1992; 92EP-0109812.
XX
XX
XX 11-JUN-1991; 91JP-0139268.
XX 12-JUL-1991; 91JP-0172794.
XX 07-OCT-1991; 91JP-0287008.
XX 16-DEC-1991; 91JP-0332329.
XX 20-APR-1992; 92JP-0099957.
XX
XX
XX (MITU ) MITSUBISHI KASEI CORP.
XX
XX
XX Hayashi N, Honda Y, Murakami T, Seki M, Takahashi K;
XX PI Teranishi Y;
XX
XX
XX WPI; 1992-417213/51.
XX DR P-PSDB; AAR29853.
XX
XX
XX New hepatitis C virus gene and its encoded protein - used for
XX diagnosing and vaccinating against hepatitis C virus infections
XX
XX
XX Disclosure; Page 149-51; 305pp; English.
XX
XX
XX The sequences given in AA032483-501 are various clones which were used
XX in the isolation of the NS2-NS4 regions of the Hepatitis C Virus
XX (HCV) gene of the invention (see also AA032442 and AA032472-82). These
XX sequences were isolated from the serum of a patient suffering from
XX hepatitis C (HC). The isolated RNA sequences were converted into
XX cDNA using transcriptase in the presence of one of the primer
XX sequences given in AA032578-79. The sequences were then amplified
XX using primer pairs. The cDNA sequences isolated represent different
XX alleles of the same region of the HCV gene. Sequence comparisons of
XX these clones showed that it is possible for a patient to carry more
XX than one HCV strain at one time. See also AA032436.
XX
XX
SQ Sequence 849 BP; 146 A; 246 C; 247 G; 210 T; 0 other;
XX
XX
Query Match 2.1%; Score 20; DB 13; Length 849;
Best Local Similarity 100.0%; Pred. No. 10;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
QY 311 CAGAGAACACAGGAAGA 330
DB 125 CAGAGAACACAGGAAGA 106
XX
XX
RESULT 14
AA032479/c
ID AA032479 standard; DNA: 1280 BP.
XX
XX
XX AA032479;
XX
XX
XX 26-APR-1993 (first entry)
XX
XX
XX HCV NS2-NS4 clone MX25026.
XX
XX
XX Clone; polypeptide; NS2-NS4; Hepatitis C; Virus; HCV; serum; HC;
XX transcriptase; cDNA; primer; allele; ss.
XX
XX
XX Hepatitis C virus.
XX
XX
XX EP518313-A.
XX
XX
XX 16-DEC-1992.
XX
XX
XX 11-JUN-1992; 92EP-0109812.
XX
XX
```

```

XX 11-JUN-1991; 91JP-0139268.
PR 12-JUL-1991; 91JP-0172794.
PR 07-OCT-1991; 91JP-0287008.
PR 16-DEC-1991; 91JP-0332329.
PR 20-APR-1992; 92JP-0099957.
XX
PA (MITU ) MITSUBISHI KASEI CORP.
XX Hayashi N, Honda Y, Murakami T, Seki M, Takahashi K;
PI Teranishi Y;
XX
XX WPI: 1992-417213/51.
DR P-PSDB: AAR29848.
XX
XX New hepatitis C virus gene and its encoded protein - used for
PT diagnosing and vaccinating against hepatitis C virus infections
XX
XX Disclosure: Page 129-31; 305pp; English.
XX
XX The sequences given in AAQ32472-82 and AAQ32442 are various clones which
CC encode the NS2-NS4 regions of the Hepatitis C Virus (HCV) gene of
CC the invention. These sequences were isolated from the serum of a
CC patient suffering from hepatitis C (HC). The isolated RNA sequences
CC were converted into cDNA using transcriptase in the presence of one
CC of the primer sequences given in AAQ3253-64. The sequences were
CC then amplified using primer pairs. The cDNA sequences isolated
CC represent different alleles of the same region of the HCV gene.
CC Sequence comparisons of these clones showed that it is possible for a
CC patient to carry more than one HCV strain at one time. See also
CC AAQ32436.
XX
SQ Sequence 1280 BP; 204 A; 353 C; 357 G; 270 T; 96 other;

Query Match 2.1%; Score 20; DB 13; Length 1280;
Best Local Similarity 100.0%; Pred. No. 10;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 311 CAGAGACACACAGAGAGA 330
DB 125 CAGAGACACACAGAGAGA 106

RESULT 15
AAQ32494/C
ID AAQ32494 standard; DNA; 1280 BP.
XX
XX AAQ32494;
XX
XX 26-APR-1993 (first entry)
XX
XX HCV NS2-NS4 clone MX25026A-1.
XX
XX Clone: polypeptide; NS2-NS4; Hepatitis C; Virus; HCV; serum; HC;
XX transcriptase; cDNA; primer; allele; ss.
XX
XX Hepatitis C virus.
XX
XX
FH Key Location/Qualifiers
FT CDS 1..1278
FT CDS /*tag= a
XX
XX EP518313-A.
XX
XX 16-DEC-1992.
XX
XX 11-JUN-1992; 92EP-0109812.
XX
XX 11-JUN-1991; 91JP-0139268.
XX 12-JUL-1991; 91JP-0172794.
XX 07-OCT-1991; 91JP-0287008.
XX 16-DEC-1991; 91JP-0332329.
XX 20-APR-1992; 92JP-0099957.

```

```

XX (MITU ) MITSUBISHI KASEI CORP.
XX Hayashi N, Honda Y, Murakami T, Seki M, Takahashi K;
PI Teranishi Y;
XX
XX WPI: 1992-417213/51.
DR P-PSDB: AAR29863.
XX
XX New hepatitis C virus gene and its encoded protein - used for
PT diagnosing and vaccinating against hepatitis C virus infections
XX
XX Disclosure: Page 165-67; 305pp; English.
XX
XX The sequences given in AAQ32483-501 are various clones which were used
CC in the isolation of the NS2-NS4 regions of the Hepatitis C Virus
CC (HCV) gene of the invention (see also AAQ32442 and AAQ32472-82). These
CC sequences were isolated from the serum of a patient suffering from
CC hepatitis C (HC). The isolated RNA sequences were converted into
CC cDNA using transcriptase in the presence of one of the primer
CC sequences given in AAQ32578-79. The sequences were then amplified
CC using primer pairs. The cDNA sequences isolated represent different
CC alleles of the same region of the HCV gene. Sequence comparisons of
CC these clones showed that it is possible for a patient to carry more
CC than one HCV strain at one time. See also AAQ32436.
XX
SQ Sequence 1280 BP; 222 A; 385 C; 376 G; 297 T; 0 other;

Query Match 2.1%; Score 20; DB 13; Length 1280;
Best Local Similarity 100.0%; Pred. No. 10;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 311 CAGAGACACACAGAGAGA 330
DB 125 CAGAGACACACAGAGAGA 106

RESULT 16
AAQ32495/C
ID AAQ32495 standard; DNA; 1280 BP.
XX
XX AAQ32495;
XX
XX 26-APR-1993 (first entry)
XX
XX HCV NS2-NS4 clone MX25026B-1.
XX
XX Clone: polypeptide; NS2-NS4; Hepatitis C; Virus; HCV; serum; HC;
XX transcriptase; cDNA; primer; allele; ss.
XX
XX Hepatitis C virus.
XX
XX
FH Key Location/Qualifiers
FT CDS 1..1278
FT CDS /*tag= a
XX
XX EP518313-A.
XX
XX 16-DEC-1992.
XX
XX 11-JUN-1992; 92EP-0109812.
XX
XX 11-JUN-1991; 91JP-0139268.
XX 12-JUL-1991; 91JP-0172794.
XX 07-OCT-1991; 91JP-0287008.
XX 16-DEC-1991; 91JP-0332329.
XX 20-APR-1992; 92JP-0099957.
XX
XX (MITU ) MITSUBISHI KASEI CORP.
XX Hayashi N, Honda Y, Murakami T, Seki M, Takahashi K;
PI Teranishi Y;
XX

```


CC	healing; (e) neurological diseases e.g. cerebral anoxia and epilepsy;
CC	and (f) infectious diseases such as viral, bacterial, fungal and
CC	parasitic infections.
CC	Note: The sequence data for this patent did not form part of the
CC	printed specification, but was obtained in electronic format directly
CC	from WIPO at http://www.int/pat/patent/patent_sequences .
XX	Sequence 1350 BP; 414 A; 250 C; 303 G; 383 T; 0 other;
SO	
Query Match	2.1%; Score 20; DB 22; Length 1350;
Best Local Similarity	100.0%; Pred. No. 10;
Matches	20; Conservative 0; Mismatches 0; Indels 0; Gaps 0.
OY	891 TCTGAGAAAACCTGGAATA 910
Db	866 TCTGAGAAAACCTGGAATA 885
RESULT 18	
ID	AAL37284
XX	AAL37284 standard; DNA; 1350 BP.
AC	
XX	AAL37284;
DT	
XX	08-JAN-2002 (first entry)
DE	
XX	Human musculoskeletal system related polynucleotide SEQ ID NO 3649.
KW	Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;
KW	antiallergic; hepatotropic; antidiabetic; antiinflammatory; antilacer;
KW	vulnerary; anticonvulsant; antibacterial; antifungal; antiparasitic;
KW	cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;
KW	neurological disease; infection; human; secreted protein;
KW	musculoskeletal system; ds.
XX	
OS	Homo sapiens.
XX	
PN	WO200155367-A1.
PD	
XX	02-AUG-2001.
PF	
XX	17-JAN-2001; 2001WO-0501338.
XX	
PR	31-JAN-2000; 2000US-0179065.
PR	04-FEB-2000; 2000US-0180628.
PR	24-FEB-2000; 2000US-0184664.
PR	02-MAR-2000; 2000US-0186350.
PR	16-MAR-2000; 2000US-0189874.
PR	17-MAR-2000; 2000US-0190076.
PR	18-APR-2000; 2000US-0198123.
PR	19-MAY-2000; 2000US-0205515.
PR	07-JUN-2000; 2000US-0209467.
PR	28-JUN-2000; 2000US-0214886.
PR	30-JUN-2000; 2000US-0215135.
PR	07-JUL-2000; 2000US-0216647.
PR	07-JUL-2000; 2000US-0216860.
PR	11-JUL-2000; 2000US-0217487.
PR	11-JUL-2000; 2000US-0217496.
PR	14-JUL-2000; 2000US-0218290.
PR	26-JUL-2000; 2000US-0220963.
PR	26-JUL-2000; 2000US-0220964.
PR	14-AUG-2000; 2000US-0224518.
PR	14-AUG-2000; 2000US-0224519.
PR	14-AUG-2000; 2000US-0225213.
PR	14-AUG-2000; 2000US-0225214.
PR	14-AUG-2000; 2000US-0225267.
PR	14-AUG-2000; 2000US-0225268.
PR	14-AUG-2000; 2000US-0225270.
PR	14-AUG-2000; 2000US-0225447.
PR	14-AUG-2000; 2000US-0225757.
PR	14-AUG-2000; 2000US-0225758.
PR	14-AUG-2000; 2000US-0225759.

PR 18-AUG-2000; 2000US-0226279.
 PR 22-AUG-2000; 2000US-0226681.
 PR 22-AUG-2000; 2000US-0226686.
 PR 22-AUG-2000; 2000US-0227182.
 PR 23-AUG-2000; 2000US-0227009.
 PR 30-AUG-2000; 2000US-0228924.
 PR 01-SEP-2000; 2000US-0229287.
 PR 01-SEP-2000; 2000US-0229343.
 PR 01-SEP-2000; 2000US-0229344.
 PR 01-SEP-2000; 2000US-0229345.
 PR 05-SEP-2000; 2000US-0229509.
 PR 05-SEP-2000; 2000US-0229513.
 PR 06-SEP-2000; 2000US-0230437.
 PR 06-SEP-2000; 2000US-0230438.
 PR 08-SEP-2000; 2000US-0231242.
 PR 08-SEP-2000; 2000US-0231243.
 PR 08-SEP-2000; 2000US-0231244.
 PR 08-SEP-2000; 2000US-0231413.
 PR 08-SEP-2000; 2000US-0231414.
 PR 08-SEP-2000; 2000US-0232080.
 PR 12-SEP-2000; 2000US-0231968.
 PR 14-SEP-2000; 2000US-0232397.
 PR 14-SEP-2000; 2000US-0232398.
 PR 14-SEP-2000; 2000US-0232399.
 PR 14-SEP-2000; 2000US-0232400.
 PR 14-SEP-2000; 2000US-0232401.
 PR 14-SEP-2000; 2000US-0233063.
 PR 14-SEP-2000; 2000US-0233064.
 PR 14-SEP-2000; 2000US-0233065.
 PR 21-SEP-2000; 2000US-0234423.
 PR 21-SEP-2000; 2000US-0234424.
 PR 25-SEP-2000; 2000US-0234997.
 PR 25-SEP-2000; 2000US-0234998.
 PR 26-SEP-2000; 2000US-0235484.
 PR 27-SEP-2000; 2000US-0235834.
 PR 27-SEP-2000; 2000US-0235836.
 PR 29-SEP-2000; 2000US-0236327.
 PR 29-SEP-2000; 2000US-0236367.
 PR 29-SEP-2000; 2000US-0236368.
 PR 29-SEP-2000; 2000US-0236369.
 PR 29-SEP-2000; 2000US-0236370.
 PR 02-OCT-2000; 2000US-0236802.
 PR 02-OCT-2000; 2000US-0237037.
 PR 02-OCT-2000; 2000US-0237038.
 PR 02-OCT-2000; 2000US-0237039.
 PR 02-OCT-2000; 2000US-0237040.
 PR 13-OCT-2000; 2000US-0239935.
 PR 13-OCT-2000; 2000US-0239937.
 PR 20-OCT-2000; 2000US-0240960.
 PR 20-OCT-2000; 2000US-0241221.
 PR 20-OCT-2000; 2000US-0241785.
 PR 20-OCT-2000; 2000US-0241786.
 PR 20-OCT-2000; 2000US-0241787.
 PR 20-OCT-2000; 2000US-0241808.
 PR 20-OCT-2000; 2000US-0241809.
 PR 20-OCT-2000; 2000US-0241826.
 PR 01-NOV-2000; 2000US-0244617.
 PR 08-NOV-2000; 2000US-0244674.
 PR 08-NOV-2000; 2000US-0246475.
 PR 08-NOV-2000; 2000US-0246476.
 PR 08-NOV-2000; 2000US-0246477.
 PR 08-NOV-2000; 2000US-0246478.
 PR 08-NOV-2000; 2000US-0246523.
 PR 08-NOV-2000; 2000US-0246524.
 PR 08-NOV-2000; 2000US-0246525.
 PR 08-NOV-2000; 2000US-0246526.
 PR 08-NOV-2000; 2000US-0246527.
 PR 08-NOV-2000; 2000US-0246528.
 PR 08-NOV-2000; 2000US-0246532.
 PR 08-NOV-2000; 2000US-0246609.
 PR 08-NOV-2000; 2000US-0246610.
 PR 08-NOV-2000; 2000US-0246611.

PR 08-NOV-2000; 2000US-0246613.
 PR 17-NOV-2000; 2000US-0249207.
 PR 17-NOV-2000; 2000US-0249208.
 PR 17-NOV-2000; 2000US-0249209.
 PR 17-NOV-2000; 2000US-0249210.
 PR 17-NOV-2000; 2000US-0249211.
 PR 17-NOV-2000; 2000US-0249212.
 PR 17-NOV-2000; 2000US-0249213.
 PR 17-NOV-2000; 2000US-0249214.
 PR 17-NOV-2000; 2000US-0249215.
 PR 17-NOV-2000; 2000US-0249216.
 PR 17-NOV-2000; 2000US-0249217.
 PR 17-NOV-2000; 2000US-0249218.
 PR 17-NOV-2000; 2000US-0249244.
 PR 17-NOV-2000; 2000US-0249245.
 PR 17-NOV-2000; 2000US-0249264.
 PR 17-NOV-2000; 2000US-0249265.
 PR 17-NOV-2000; 2000US-0249297.
 PR 17-NOV-2000; 2000US-0249299.
 PR 17-NOV-2000; 2000US-0249300.
 PR 01-DEC-2000; 2000US-0250160.
 PR 01-DEC-2000; 2000US-0250391.
 PR 05-DEC-2000; 2000US-0251030.
 PR 05-DEC-2000; 2000US-0251988.
 PR 05-DEC-2000; 2000US-0256719.
 PR 08-DEC-2000; 2000US-0251479.
 PR 08-DEC-2000; 2000US-0251856.
 PR 08-DEC-2000; 2000US-0251868.
 PR 08-DEC-2000; 2000US-0251869.
 PR 08-DEC-2000; 2000US-0251989.
 PR 11-DEC-2000; 2000US-0251990.
 PR 05-JAN-2001; 2001US-0259678.
 (HUMA-) HUMAN GENOME SCI INC.
 Rosen CA, Barash SC, Ruben SM;
 WPI; 2001-451937/48.
 Isolated polypeptide for treating, preventing and/or prognosing
 disorders related to the musculoskeletal system including
 musculoskeletal cancers and also for testing and detection e.g.
 diagnosis -
 XX
 PS Example 2; SEQ ID NO 3649; 781pp + Sequence Listing; English.
 CC The invention relates to novel genes (AAL34669-AAL37666) and proteins
 CC (ABR03087-ABR04109) associated with the musculoskeletal system useful
 CC for preventing, treating or ameliorating medical conditions e.g. by
 CC protein or gene therapy. The genes are isolated from a range of human
 CC tissues disclosed in the specification. The nucleic acids, proteins,
 CC antibodies and (ant)agonists are useful in the diagnosis, treatment
 CC and prevention of: (a) cancer, e.g. breast and ovarian cancer and
 CC other cancers of the adrenal gland, bone, bone marrow, breast,
 CC gastrointestinal tract, liver, lung, or urogenital; (b) immune
 CC disorders e.g. Addison's disease, allergies, autoimmune hemolytic
 CC anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's disease,
 CC multiple sclerosis, rheumatoid arthritis and ulcerative colitis;
 CC (c) cardiovascular disorders such as myocardial ischaemias; (d) wound
 CC healing; (e) neurological diseases e.g. cerebral anoxia and epilepsy;
 CC and (f) infectious diseases such as viral, bacterial, fungal and
 CC parasitic infections.
 CC Note: The sequence data for this patent did not form part of the
 CC printed specification, but was obtained in electronic format directly
 CC from WPI at ftp.wipo.int/pub/published_pat_sequences.
 CC
 XX Sequence 1350 BP; 414 A; 250 C; 303 G; 383 T; 0 other;
 Query Match 2.1%; Score 20; DB 22; Length 1350;
 Best Local Similarity 100.0%; Pred. No. 10;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	891	TCTGAGAAAACCTGGATA	910
Db	866	TCTGAGAAAACCTGGATA	865
	RESULT 19		
	AAAL37285		
	AAAL37285 standard; DNA; 1350 BP.		
XX	AAAL37285;		
AC			
XX			
DT	08-JAN-2002 (first entry)		
DE	Human musculoskeletal system related polynucleotide SEQ ID NO 3650.		
XX			
KW	Cytostatic; immunosuppressive; neurotropic; neuroprotective; antiviral;		
KW	antiallergic; hepatotropic; antidiabetic; antiinflammatory; antitumor;		
KW	vulnerary; anticonvulsant; antibacterial; antifungal; antiparasitic;		
KW	cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder		
KW	neurological disease; infection; human; secreted protein;		
KW	musculoskeletal system; ds.		
XX			
OS	Homo sapiens.		
XX			
PN	WO200155367-A1.		
XX			
PD			
XX			
PF	02-AUG-2001.		
	17-JAN-2001; 2001MO-US01338.		
XX			
PR	31-JAN-2000; 2000US-0179065.		
PR	04-FEB-2000; 2000US-0180628.		
PR	24-FEB-2000; 2000US-0184664.		
PR	02-MAR-2000; 2000US-0186350.		
PR	16-MAR-2000; 2000US-0189874.		
PR	17-MAR-2000; 2000US-0190076.		
PR	18-APR-2000; 2000US-0198123.		
PR	19-MAY-2000; 2000US-0205515.		
PR	07-JUN-2000; 2000US-0209467.		
PR	28-JUN-2000; 2000US-0214886.		
PR	30-JUN-2000; 2000US-0215135.		
PR	07-JUL-2000; 2000US-0216647.		
PR	07-JUL-2000; 2000US-0216880.		
PR	11-JUL-2000; 2000US-0217487.		
PR	11-JUL-2000; 2000US-0217496.		
PR	14-JUL-2000; 2000US-0218290.		
PR	26-JUL-2000; 2000US-0220963.		
PR	26-JUL-2000; 2000US-0220964.		
PR	14-AUG-2000; 2000US-0224516.		
PR	14-AUG-2000; 2000US-0224519.		
PR	14-AUG-2000; 2000US-0225213.		
PR	14-AUG-2000; 2000US-0225214.		
PR	14-AUG-2000; 2000US-0225266.		
PR	14-AUG-2000; 2000US-0225267.		
PR	14-AUG-2000; 2000US-0225268.		
PR	14-AUG-2000; 2000US-0225270.		
PR	14-AUG-2000; 2000US-0225547.		
PR	14-AUG-2000; 2000US-0225757.		
PR	14-AUG-2000; 2000US-0225758.		
PR	14-AUG-2000; 2000US-0225759.		
PR	18-AUG-2000; 2000US-0226279.		
PR	22-AUG-2000; 2000US-0226681.		
PR	22-AUG-2000; 2000US-0226866.		
PR	22-AUG-2000; 2000US-0227182.		
PR	23-AUG-2000; 2000US-0227009.		
PR	30-AUG-2000; 2000US-0228924.		
PR	01-SEP-2000; 2000US-0229287.		
PR	01-SEP-2000; 2000US-0229343.		
PR	01-SEP-2000; 2000US-0229344.		
PR	01-SEP-2000; 2000US-0229345.		
PR	05-SEP-2000; 2000US-0229509.		
PR	05-SEP-2000; 2000US-0229513.		
PR	06-SEP-2000; 2000US-0230437.		

```

PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
PA
XX Rosen CA, Barash SC, Ruben SM;
PI
XX WPI: 2001-451937/48.
XX
XX Isolated polypeptide for treating, preventing and/or prognosing
PT disorders related to the musculoskeletal system including
PT musculoskeletal cancers and also for testing and detection e.g.
PT diagnosis.
XX
XX Example 2: SEQ ID NO 3650; 781bp + Sequence Listing; English.
XX
CC The invention relates to novel genes (AA034669-AA037666) and proteins
CC (AB03087-AB04109) associated with the musculoskeletal system useful
CC for preventing, treating or ameliorating medical conditions e.g. by
CC protein or gene therapy. The genes are isolated from a range of human
CC tissues disclosed in the specification. The nucleic acids, proteins,
CC antibodies and (ant)agonists are useful in the diagnosis, treatment
CC and prevention of: (a) cancer, e.g. breast and ovarian cancer and
CC other cancers of the adrenal gland, bone, bone marrow, breast,
CC gastrointestinal tract, liver, lung, or urogenital; (b) immune
CC disorders e.g. Addison's disease, allergies, autoimmune haemolytic
CC anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's disease,
CC multiple sclerosis, rheumatoid arthritis and ulcerative colitis;
CC (c) cardiovascular disorders such as myocardial ischaemias; (d) wound
CC healing; (e) neurological diseases e.g. cerebral anoxia and epilepsy;
CC and (f) infectious diseases such as viral, bacterial, fungal and
CC parasitic infections.
CC Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 1350 BP; 413 A; 250 C; 303 G; 384 T; 0 other;
SQ
Query Match 2.1%; Score 20; DB 22; Length 1350;
Best Local Similarity 100.0%; Pred. No. 10;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 891 TCTGAGAAAACCTGGAATA 910
DB 866 TCTGAGAAAACCTGGAATA 885

```

```

RESULT 20
AA032442/c
ID AA032442 standard; DNA; 3564 BP.
XX
XX AA032442;
AC
XX
XX 26-APR-1993 (first entry)
XX
XX

```

```

DE HCV NS2-NS4 clone MX25N15.
XX
XX Clone; polypeptide; NS2-NS4; Hepatitis C; Virus; HCV; serum; HC;
KM transcriptase; cDNA; primer; allele; ss.
XX
XX Hepatitis C virus.
OS
XX
XX EP518313-A.
PN
XX
XX 16-DEC-1992.
PD
XX
XX 11-JUN-1992; 92EP-0109812.
PF
XX
XX 11-JUN-1991; 91JP-0139268.
PR 12-JUL-1991; 91JP-0172794.
PR 07-OCT-1991; 91JP-0287008.
PR 16-DEC-1991; 91JP-0332329.
PR 20-APR-1992; 92JP-0099957.
XX
XX (MITU) MITSUBISHI KASEI CORP.
PA
XX
XX Hayashi N, Honda Y, Murakami T, Seki M, Takahashi K;
PI Teranishi Y;
XX
XX WPI: 1992-417213/51.
DR P-PSDB: AAR29660.
XX
XX New hepatitis C virus gene and its encoded protein - used for
PT diagnosing and vaccinating against hepatitis C virus infections
PT
XX Disclosure; Page 139-46; 3055bp; English.
XX
XX The sequences given in AA03472-82 and AA032442 are various clones which
XX encode the NS2-NS4 regions of the Hepatitis C virus (HCV) gene of
XX the invention. These sequences were isolated from the serum of a
XX patient suffering from Hepatitis C (HC). The isolated RNA sequences
XX were converted into cDNA using transcriptase in the presence of one
XX of the primer sequences given in AA032553-64. The sequences were
XX then amplified using primer pairs. The cDNA sequences isolated
XX represent different alleles of the same region of the HCV gene.
XX Sequence comparisons of these clones showed that it is possible for a
XX patient to carry more than one HCV strain at one time. See also
XX AA032436.
XX
XX Sequence 3564 BP; 640 A; 1019 C; 1026 G; 747 T; 132 other;
SQ
Query Match 2.1%; Score 20; DB 13; Length 3564;
Best Local Similarity 100.0%; Pred. No. 9.8;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 311 CAGACAGACACAGAGAAAG 330
DB 125 CAGACAGACACAGAGAAAG 106

```

```

RESULT 21
AA032501/c
ID AA032501 standard; DNA; 3564 BP.
XX
XX AA032501;
AC
XX
XX 26-APR-1993 (first entry)
DT
XX
XX HCV NS2-NS4 clone N25N15-1.
DE
XX
XX Clone; polypeptide; NS2-NS4; Hepatitis C; Virus; HCV; serum; HC;
KM transcriptase; cDNA; primer; allele; ss.
XX
XX Hepatitis C virus.
OS
XX
XX EP518313-A.
PN
XX
XX 16-DEC-1992.
PD

```

```

XX 11-JUN-1992; 92EP-0109812.
PF 11-JUN-1991; 91JP-0139268.
PR 12-JUL-1991; 91JP-0172794.
PR 07-OCT-1991; 91JP-0287008.
PR 16-DEC-1991; 91JP-0332329.
PR 20-APR-1992; 92JP-0099957.
XX
PA (MITU ) MITSUBISHI KASEI CORP.
PI Hayashi N, Honda Y, Murakami T, Seki M, Takahashi K;
PI Teranishi Y;
DR WPI: 1992-417213/51.
DR P-PSDB; AAR29870.
XX
PT New hepatitis C virus gene and its encoded protein - used for
PT diagnosing and vaccinating against hepatitis C virus infections
PS
XX Disclosure; Page 186-92; 305pp; English.
XX
CC The sequences given in AAQ32483-501 are various clones which were used
CC in the isolation of the NS2-NS4 regions of the Hepatitis C Virus
CC (HCV) gene of the invention (see also AAQ32442 and AAQ32472-82). These
CC sequences were isolated from the serum of a patient suffering from
CC hepatitis C (HC). The isolated RNA sequences were converted into
CC cDNA using transcriptase in the presence of one of the primer
CC sequences given in AAQ32578-79. The sequences were then amplified
CC using primer pairs. The cDNA sequences isolated represent different
CC alleles of the same region of the HCV gene. Sequence comparisons of
CC these clones showed that it is possible for a patient to carry more
CC than one HCV strain at one time. See also AAQ32436.
XX
SQ Sequence 3564 BP; 668 A; 1064 C; 1046 G; 786 T; 0 other;
Query Match 2.1%; Score 20; DB 13; Length 3564;
Best Local Similarity 100.0%; Pred. No. 9.8;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 311 CAGAGAACACACAGAGAAGA 330
DB 125 CAGAGAACACACAGAGAAGA 106
RESULT 22
ABL28818/C
ID ABL28818 standard; DNA; 4235 BP.
XX
AC ABL28818;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster genomic polynucleotide SEQ ID NO 37927.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical; gene; ds.
OS Drosophila melanogaster.
PN WO200171042-A2.
XX
PD 27-SEP-2001.
PF 23-MAR-2001; 2001WO-US09231.
PR 23-MAR-2000; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150.
XX
PA (PEKE ) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
PT

```

```

DR WPI: 2001-656860/75.
XX
XX New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
PT
XX Claim 1; SEQ ID NO 37927; 21pp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (AB161176-AB130511), expressed DNA
CC sequences (AB101840-AB161175) and the encoded proteins
CC (ABBS7737-ABB72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pft_sequences.
XX
SQ Sequence 4235 BP; 1103 A; 850 C; 759 G; 1523 T; 0 other;
Query Match 2.1%; Score 20; DB 23; Length 4235;
Best Local Similarity 100.0%; Pred. No. 9.8;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 189 TCCTCCTTATTAATCATGCG 208
DB 3940 TCCTCCTTATTAATCATGCG 3921
RESULT 23
AAQ32436/C
ID AAQ32436 standard; DNA; 7911 BP.
XX
AC AAQ32436;
XX
DT 26-APR-1993 (first entry)
XX
DE HCV antigen clone T7N1-30.
XX
KW Clone; Hepatitis C Virus; HCV; core-envelope; NS1(gp70); NS2-NS4;
KW NS4-NS5; region; diagnostic method; antibody; suppress; control;
KW proteolytic; process; precursor; polypeptide; ss.
XX
OS Hepatitis C virus.
XX
FH Key Location/Qualifiers
FT CDS 369..7898
FT /tag= a
XX
PN EP518313-A.
XX
PD 16-DEC-1992.
XX
PE 11-JUN-1992; 92EP-0109812.
XX
PR 11-JUN-1991; 91JP-0139268.
PR 12-JUL-1991; 91JP-0172794.
PR 07-OCT-1991; 91JP-0287008.
PR 16-DEC-1991; 91JP-0332329.
PR 20-APR-1992; 92JP-0099957.
XX
PA (MITU ) MITSUBISHI KASEI CORP.
XX
PI Hayashi N, Honda Y, Murakami T, Seki M, Takahashi K;
PI Teranishi Y;
DR WPI: 1992-417213/51.
DR P-PSDB; AAR29527.
XX
PT New hepatitis C virus gene and its encoded protein - used for
PT diagnosing and vaccinating against hepatitis C virus infections

```

XX Claim 1 and 3; Page 259-272; 305pp; English.
PS
XX
CC The sequences given in AA032436-539 are various clones which were all
CC derived from the isolated Hepatitis C Virus (HCV) gene of the
CC invention. AA032436 contains the entire protein gene, whereas AA032444-
CC 53 and AA032438-39, AA032454-68, AA032472-82 and AA032442, and
CC AA032502-12 and AA032443 encode the core-envelope, NS1(gp70), NS2-NS4
CC and NS4-NS5 regions respectively. AA032436 represents the entire gene
CC sequence. The HCV gene is useful in the development of a diagnostic
CC method which is more accurate and effective than conventional ones, in
CC the detection of antibodies raised against a wide range of HCVs which
CC have been hardly detected before. The complete gene may be used in an
CC in vitro screening system for a substance capable of specifically
CC suppressing or controlling a proteolytic processing of a precursor
CC polypeptide of HCV.
XX
SQ Sequence 7911 BP; 1556 A; 2363 C; 2292 G; 1700 T; 0 other;
Query Match 2.1%; Score 20; DB 13; Length 7911;
Best Local Similarity 100.0%; Pred. No. 9.7;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 311 CAGAGACACACAGGAAGA 330
|||
DB 2686 CAGAGACACACAGGAAGA 2667

Search completed: April 26, 2003, 11:33:26
Job time : 211 secs

GenCore version 5.1.4-p5-4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 26, 2003, 10:58:07 ; Search time 1108 Seconds

(without alignments)
13944.489 Million cell updates/sec

Title: US-09-880-729a-1

Perfect score: 954

Sequence: 1 atgggtgtgacgttcttga.....gagagatagcatgataaa 954

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 16154066 seqs, 8097743376 residues

Word size : 20

Total number of hits satisfying chosen parameters: 18

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database :

EST:*
1: em_estba:*
2: em_estham:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estp1:*
7: em_estro:*
8: em_hlc:*
9: gb_est1:*
10: gb_est2:*
11: gb_hlc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	23	2.4	563	10	AM308797 s72a11.y
2	21	2.2	252	10	AV359581 AV359581
3	20	2.1	139	9	AV097171 AV097171
4	20	2.1	258	10	BB153117 BB153117
5	20	2.1	307	9	AA312956 EST183735
6	20	2.1	385	9	AI592815 mt70b05.y

Result	LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS	SOURCE	ORGANISM	REFERENCE	AUTHORS	TITLE	JOURNAL	COMMENT
7	AM308797	563 bp mRNA linear EST 02-DEC-2001 s72a11.y1 Gm-cl013 Glycine max cDNA clone GENOME SYSTEMS CLONE ID: Gm-cl013-5109 5' similar to SW:CAHC_PEA P17067 CARBONIC ANHYDRASE, CHLOROPLAST PRECURSOR ;, mRNA sequence.	AM308797	AM308797.1	GI:6724398	soybean.	Glycine max	1 (bases 1 to 563)	Shoemaker,R., Kelm,P., Vodkin,L., Erpeliding,J., Corryell,V., Khanna,A., Bolla,B., Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Ralster,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R., Waterston,R. and Wilson,R.	Public Soybean EST Project	Unpublished (1999)	Contact: Shoemaker R/Public Soybean EST Project
8	AM308797	563 bp mRNA linear EST 02-DEC-2001 s72a11.y1 Gm-cl013 Glycine max cDNA clone GENOME SYSTEMS CLONE ID: Gm-cl013-5109 5' similar to SW:CAHC_PEA P17067 CARBONIC ANHYDRASE, CHLOROPLAST PRECURSOR ;, mRNA sequence.	AM308797	AM308797.1	GI:6724398	soybean.	Glycine max	1 (bases 1 to 563)	Shoemaker,R., Kelm,P., Vodkin,L., Erpeliding,J., Corryell,V., Khanna,A., Bolla,B., Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Ralster,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R., Waterston,R. and Wilson,R.	Public Soybean EST Project	Unpublished (1999)	Contact: Shoemaker R/Public Soybean EST Project

ALIGNMENTS

FEATURES

source

1. 563

/organism="Glycine max"

/db_xref="taxon:3847"

/clone="GENOME SYSTEMS CLONE ID: Gm-cl013-5109"

/clone_11b="Gm-cl013"

/tissue_type="Whole seedlings, 2-3 week old seedlings, greenhouse grown"

/lab_host="XL10-Gold"

/note="Vector: pBluescript II XR; Site:1: EcoRI; Site:2: XhoI. This cDNA library was constructed from mRNA isolated from whole seedlings of 2-3 week old greenhouse grown plants. The cDNA library was prepared using the Stratagene pBluescript II XR cDNA library construction kit. Complementary DNA was synthesized from mRNA using a primer consisting of a poly (dI) sequence with a blunt-ended cDNA site. EcoRI adapters were ligated to the blunt-ended cDNA fragments followed by XhoI digestion. The cDNA fragments were directionally cloned into the EcoRI-XhoI restriction site of the pBluescript vector. The ligated cDNA fragments


```

/clone="2410026A20"
/clone_lib="Mus musculus C57BL/6J ES cell"
/cell_type="ES cell"
/note="organ: mammary gland; Vector: p773D-Pac (Pharmacia)
with a modified polylinker. Site_1: Not I; Site_2: Eco
RI; 1st strand cDNA was primed with a Not I - oligo(dT)
primer [5']
TGTTCACATCTGAGTGGAGCGCGCGCATGCTTTTCTTTTCTTTTCTTTT
T 3']; double-stranded cDNA was ligated to Eco RI
adaptors (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of the modified p773 vector.
RNA provided by Dr. Minoru Ko, Wayne State Univ. Library
constructed and normalized by Bento Soares and M.Fatima
Bonaldo."
BASE COUNT      24 a      26 c      34 g      55 t
ORIGIN
Query Match      2.1%; Score 20; DB 9; Length 139;
Best Local Similarity 100.0%; Pred. No. 25;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 726 AAAAAACAACAAAGACCA 745
Db 100 AAAAAACAACAAAGACCA 81

RESULT 4
BBI53117/c      258 bp      mRNA      linear      EST 28-JUN-2000
BBI53117 RIKEN full-length enriched, 16 days neonate thymus Mus
musculus cDNA clone A130010K07 3' similar to U88328 Mus musculus
suppressor of cytokine signalling-3 (SOCS-3) mRNA, mRNA sequence.
BBI53117
BBI53117.1 GI:8808054
EST
house mouse.
Mus musculus
SOURCE
ORGANISM
Mammalia: Eutheria: Rodentia: Sciurognathi: Muridae: Murinae: Mus.
1 (bases 1 to 258)
Kono, H., Aizawa, K., Akahira, S., Akiyama, J., Arakawa, T., Carninci,
P., Endo, T., Fukuda, S., Fukunishi, Y., Hara, A., Hayatsu, N.,
Hizzone, T., Hori, F., Ishii, Y., Ishikawa, J., Ishikawa, T., Itoh, M.,
Izawa, M., Kadota, K., Kagawa, I., Kai, C., Kawai, J., Kikuchi, N.,
Kiyosawa, H., Kojima, Y., Kondo, S., Koyas, S., Kurihara, C., Kusabe, M.,
Matsuyama, T., Miki, R., Mizuno, Y., Nakamura, M., Oda, H., Okazaki, Y.,
Ono, T., Owa, C., Saito, H., Sakai, C., Sato, K., Shibata, K., Shibata,
Y., Shigemoto, Y., Shinagawa, A., Shiraki, T., Sogabe, Y., Sugahara, Y.,
Suzuki, H., Suzuki, H., Tagawa, A., Takahashi, F., Tomimaga, N., Toya,
T., Tsunoda, Y., Watahiki, A., Watanabe, S., Yamamura, T., Yamanaka, I.,
Yano, R., Yasunishi, A., Yokota, T., Yoshida, K., Yoshiki, A., Yoshino,
M., Muramatsu, M. and Hayashizaki, Y.
RIKEN Mouse ESTs (Kono, H., et al.)
Unpublished (2000)
Contact: Yoshinori Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic
Sciences Center(GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@sc.riken.go.jp,
URL: http://genome-gsc.riken.go.jp/
Carninci, P., Nishiyama, Y., Westover, A., Itoh, M., Nagaoaka, S., Sasaki,
N., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
Thermolabile and thermocatalytic of thermolabile enzymes by
trehalose and its application for the synthesis of full length
cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)
Itoh, M., Katsunari, T., Akiyama, J., Shibata, K., Izawa, M., Kawai, J.,
Tomaru, Y., Carninci, P., Shibata, Y., Ozawa, Y., Muramatsu, M., Okazaki,
Y. and Hayashizaki, Y.
Automated filtration-based high-throughput plasmid preparation
system. Genome Res. 9 (5), 463-470 (1999)

```

```

Carninci, P. and Hayashizaki, Y.
High-efficiency full-length cDNA cloning. Methods Enzymol. 303,
19-44 (1999)
Please visit our web site (http://genome.rtc.riken.go.jp) for
further details.
FEATURES
source
Location/Qualifiers
1..258
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="A130010K07"
/clone_lib="RIKEN full-length enriched, 16 days neonate
thymus"
/tissue_type="thymus"
/dev_stage="16 days neonate"
/lab_host="DH10B"
/note="Site_1: SalI; Site_2: BamHI; cDNA library was
prepared and sequenced in Mouse Genome Encyclopedia
Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in
RIKEN. Division of Experimental Animal Research in Riken
contributed to prepare mouse tissues. 1st strand cDNA was
primed with a primer [5']
GAGAGAGAGAGATCCAGAGCGCTCTTTTCTTTTCTTTTCTTTTCTTTT
3'], cDNA was
transcribed by using trehalose thermo-activated reverse
transcriptase and subsequently enriched for full-length by
cap-trapper. cDNA went through one round of normalization
to Rot = 10.0 and subtraction to Rot = 185.0. Second
strand cDNA was prepared with the primer adaptor of
sequence [5'] GAGAGAGAGATTCAGATTAATTAATTAATTCCTCCCTCC
3']. cDNA was cleaved with XhoI and BamHI. Vector: a
modified pBluescript KS(+) after bulk excision from lambda
FLC I."
BASE COUNT      77 a      48 c      37 g      95 t      1 others
ORIGIN
Query Match      2.1%; Score 20; DB 10; Length 258;
Best Local Similarity 100.0%; Pred. No. 27;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 386 TATTTTGAATTCGAAT 405
Db 188 TATTTTGAATTCGAAT 169.

RESULT 5
AA312956      307 bp      mRNA      linear      EST 19-APR-1997
DEFINITION
AA312956 Bone VII Homo sapiens cDNA 5' end, mRNA sequence.
ACCESSION
AA312956.1 GI:1965304
VERSION
AA312956.1
KEYWORDS
EST.
SOURCE
human.
ORGANISM
Homo sapiens
Mammalia: Eutheria: Primates: Catarrhini: Hominoidea: Homo.
1 (bases 1 to 307)
Adams, M.D., Kerlavage, A.R., Fleischmann, R.D., Fuldner, R.A., Bult,
C.J., Lee, N.H., Kirkness, E.F., Weinstock, K.G., Gocayne, J.D., White,
O., Sutton, G., Blake, J.A., Brannon, R.C., Wan, W.-C., Clayton, R.A.,
Cline, T.R., Cotton, M.D., Earle-Hughes, J., Fine, L.D., Fitzgerald,
L.M., Fitzhugh, W.M., Fritchman, J.L., Geoghegan, N.S., Glodak, A.,
Guelm, C.L., Hanna, M.C., Hedblom, E., Hinkle, P.S., Jr., Kelley, J.M.,
Kelley, J.C., Liu, L.-I., Marmaros, S.M., Merrick, J.M.,
Moreno-Palauques, R.F., McDonald, L.A., Nguyen, D.T., Pelligrino, S.M.,
Phillips, C.A., Ryder, S.E., Scott, J.L., Sauder, D.M., Shirley, R.,
Small, K.V., Spriggs, T.A., Uterback, T.R., Weidman, J.F., Li, Y.,
Bednarik, D.P., Cao, L., Cepeda, M.A., Coleman, T.A., Collins, E.J.,
Dimke, D., Feng, D.-F., Ferris, A., Fischer, C., Hastings, G.A., He, W.W.,
Hu, J.S., Greene, J.M., Gruber, J., Hudson, P., Kim, A.K., Kozak, D.L.,
Kunsch, C., Hungjun, J., Li, H., Meisner, P.S., Olsen, H., Raymond, L.,
Wei, Y.F., Wing, J., Xu, C., Yu, G.L., Ruben, S.M., Dillon, P.J., Fannon,
M.R., Rosen, C.A., Haseltine, W.A., Fields, C., Fraser, C.M. and
Venter, J.C.

```

TITLE	Initial assessment of human gene diversity and expression patterns based upon 83 million nucleotides of cDNA sequence
JOURNAL	Nature 377 (6547 Suppl.), 3-174 (1995)
MEDLINE	96026280
COMMENT	Other ESTs: THC183992
	Database: EMBL/Genbank/DB

Initial assessment of human gene diversity and expression patterns based upon 83 million nucleotides of cDNA sequence
Nature 377 (6547 Suppl.), 3-174 (1995)

Other ESTs: THC183992
96026280

Contact: Kerlavage, AR
Bioinformatics

The Institute for Genomic Research
9712 Medical Center Drive, Rockville, MD 20850 USA
Tel: 3018699056
Fax: 3018699423

Email: arkerlav@tigr.org

For clone availability, additional sequence and expression information related to this EST, please check the TIGR Human Gene Index (<http://www.tigr.org/tdb/hgi.hml>)

Seq primer: M13 Reverse.

FEATURES	Location/Qualifiers
source	1. .307

BASE COUNT	ORIGIN
98 a	38 c 52 g 95 t 4 others

Query Match	2.1%	Score	20	DB	9	Length	307
Best Local Similarity	100.0%	Pred. No.	27				
Matches	20	Conservative	0	Mismatches	0	Indels	0
						Gaps	0

QY 891 TCTGAGAAAAACCTGGAATA 910
|||||
Db 111 TCTGAGAAAAACCTGGAATA 130

RESULT	6
AIS92815/c	
LOCUS	385 bp mRNA linear EST_15-MAR-2000
DEFINITION	m170b05.y1 Soares mouse lymph node NBLM Mus musculus cDNA clone
IMAGE:	G35217 5' , mRNA sequence.
ACCESSION	A1592815
VERSION	A1592815.1
KEYWORDS	GI:4601863
SOURCE	Est.
	house mouse.

ORGANISM	REFERENCE	AUTHORS
<i>Mus musculus</i>	1 (bases 1 to 385)	Marrta, M., Hillier, L., Kucaba, T., Martin, J., Beck, C., Wyjcie, T., ...
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus		

TITLE
JOURNAL
COMMENT

The WashU-NCI Mouse EST Project 1999
 Unpublished (1999)
 Contact: Marra M/WashU-NCI Mouse EST Project 1999
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: mouseest@watson.wustl.edu
 This clone is available royalty-free through LINL ; contact the
 IMAGE Consortium (infoimage.llnl.gov) for further information.
 This read is a RESSEQUENCE of a previously sequenced mouse clone
 This read has been verified (found to hit its original self in the
 correct orientation)
 putative full length read
 vector to vector length is 391
 MGI:387209
 Seq primer: -40RP from Gibco
 high quality sequence stop: 371

POLYA=No.	Location/Qualifiers	FEATURES	source
	1. .385		

```

location/Qualifiers
1..385
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone_image:635217"
/clone_1ib="Soares mouse lymph node NbM1N"
/sex="male"
/tissue_type="lymph node"
/dev_stage="4 weeks"
/lab_host="DH10B"
/note="Organ: lymph node; Vector: pT73D-Pac (Pharmacia)
with a modified polylinker; Site_1: Not I; Site_2: Eco RI;
1st strand cDNA was primed with a Not I - oligo(dT) primer"
15'
GGTACCAATCTGAGGAGGCGCGGATACCTTTTTTTTTTTTTTTTTTTTTT
3'] double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT73 vector. RNA
provided by Dr. Bertrand Jordan. Library constructed and
normalized by Bento Soares and M.Fátima Bonaldo."

```

BASE COUNT	99 a	79 c	76 g	131 t
ORIGIN				
Query Match	2.1%;	Score 20;	DB 9;	Length 385;
Best Local Similarity	100.0%;	Pred. No. 28;		
Matches 20;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0

QY 726 AAAAAAGAACAAAAGACCAA 745
|||||
Db 46 AAAAAAGAACAAAAGACCAC 27

RESULT 7	486 bp	DNA	linear	GSS 22-JUN-1999
LOCUS	A0653914			
DEFINITION	Sheared DNA-BD15, TR Sheared DNA Trypanosoma brucei genomic clone			
ACCESSION	A0653914			
VERSION	A0653914.1	GI:5147100		
KEYWORDS	GSS.			
SOURCE	Trypanosoma brucei.			
ORGANISM	Trypanosoma brucei			

REFERENCE	AUTHORS	TITLE	JOURNAL	COMMENT
1 (bases 1 to 486)	El-Sayed, N., Zhao, S., Zhao, H., Gill, S., Suh, E., Malek, J., Fujii, C., Gerrard, C., Ieech, V., de Jong, P., Ullu, E., Melville, S., Donelson, J., Fraser, C. and Adams, M.	Determination of clone end sequences from Trypanosoma brucei GUTat 10.1 sheared DNA library	Unpublished (1999)	Other_GSSS: Sheared DNA-8DJ5.TF

Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: nelsayed@tigr.org
Clones are derived from the *Trypanosoma brucei* GUTat 10.1 sheared
DNA library constructed at tigr. Clones will be available for
distribution through ARCC. Sheared DNA end sequences search page:
<http://www.tigr.org/tdb/mbd/ldbdb/>.
Seq primer: M13-Reverse

FEATURES	location/qualifiers
source	1. 486
	/organism="Trypanosoma brucei"
	/strain="TREU927/4 GUTat 10.1"
	/db_xref="taxon:5691"
	/clone="Sheared DNA-8D15"

/clone_lib="Sheared DNA"
 /note="Vector: pUC18; Site 1: SmaI; Constructed at The Institute for Genomic Research (TIGR), Rockville, MD. Genomic DNA isolated from a cloned population of Trypanosoma brucei (TREU927/4 CUTat 10.1) was mechanically sheared to give a tight size distribution (approx 2 kb). The v + 1 method used for the library construction is described in detail in Smith, H.O. and Venter, J.C. (Making small insert libraries for whole genome shotgun sequencing projects. In Genome Sequencing: A Practical Approach, eds. M. Vaudin and B. Barrell, Oxford University Press, 1999)."
 BASE COUNT 93 a 115 c 128 g 150 t
 ORIGIN
 Query Match 2.1%; Score 20; DB 17; Length 486;
 Best Local Similarity 100.0%; Pred. No. 28;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 681 TCAGAAACATTGTATAGAG 700
 |||
 Db 391 TCAGAAACATTGTATAGAG 410

RESULT 8
 BG081375 509 bp mRNA linear EST 26-JAN-2001
 LOCUS H3064D12-5 NIA Mouse 15K cDNA Clone Set Mus musculus cDNA clone
 DEFINITION H3064D12-5', mRNA sequence.
 ACCESSION BG081375
 VERSION BG081375.1 GI:12563943
 KEYWORDS EST.
 SOURCE house mouse.
 ORGANISM Mus musculus.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 509)
 Kargul, G.J., Dudekula, D.B., Qian, Y., Lim, M.K., Jaradat, S.A., Tanaka, T.S., Carter, M.G. and Ko, M.S.H.
 Verification and initial annotation of NIA mouse 15K cDNA clone set unpublished (2001)
 Other_FEATURES: H3064D12-3
 Contact: George J. Kargul
 Laboratory of Genetics
 National Institute on Aging/National Institutes of Health
 333 Cassell Drive, Suite 4000, Baltimore, MD 21224-6820, USA
 Email: cdnagisun.grc.nia.nih.gov
 This clone set has been freely distributed to the community. Please visit <http://lgsun.grc.nia.nih.gov/cDNA/15k.html> for details.
 Plate: H3064 row: D column: 12
 Seq primer: -21M13 Reverse
 High quality sequence stop: 509
 POLY-A-No.

FEATURES
 source
 1. 509
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="niaesr:H3064D12-5"
 /db_xref="taxon:10090"
 /clone="H3064D12"
 /clone_lib="NIA Mouse 15K cDNA Clone Set"
 /sex="Clones arrayed from a variety of cDNA libraries"
 /dev_stage="Clones arrayed from a variety of cDNA libraries"
 /lab_host="DH10B"
 /note="Vector: pSPORT1; Site 1: SalI; Site 2: NotI; This clone is among a rearranged set of 15,247 clones from 11 embryo cDNA libraries (including preimplantation stage embryos from unfertilized egg to blastocyst, embryonic part of E7.5 embryos, extraembryonic part of E7.5 embryos, and E12.5 female mesonephros/gonad) and one newborn ovary cDNA library. Average insert size 1.5 kb. All source libraries are cloned unidirectionally with Oligo(dT

)-Not primers. References include: (1) Genome-wide expression profiling of mid-gestation placenta and embryo using a 15,000 mouse developmental cDNA microarray, 2000, Proc. Natl. Acad. Sci. U S A. 97: 9127-9132. (2) Large-scale cDNA analysis reveals phased gene expression patterns during preimplantation mouse development, 2000, Development, 127: 1737-1749; (3) Genome-wide mapping of unselected transcripts from extraembryonic tissue of 7.5-day mouse embryos reveals enrichment in the t-complex and under-representation on the X chromosome, 1998, Hum Mol Genet 7: 1967-1978."
 BASE COUNT 176 a 81 c 90 g 160 t 2 others
 ORIGIN
 Query Match 2.1%; Score 20; DB 12; Length 509;
 Best Local Similarity 100.0%; Pred. No. 28;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 336 TGCTCTTTGGAAACAATG 355
 |||
 Db 289 TGCTCTTTGGAAACAATG 308

RESULT 9
 AW963311 608 bp mRNA linear EST 01-JUN-2000
 LOCUS AW963311 MAGE resequences, MACH Homo sapiens cDNA, mRNA sequence.
 DEFINITION AW963311
 ACCESSION AW963311 GI:8153147
 VERSION AW963311.1 GI:8153147
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 608)
 Hegde, P., Qi, R., Abernathy, K., Dharap, S., Gaspar, R., Gay, C., Holt, I.E., Saeed, A.I., Sharov, V., Lee, N.H., Yeatman, T.J. and Quackenbush, J.
 Assessment of gene expression patterns in a model of colon tumor metastasis using a 19,200 element cDNA microarray unpublished (2000)
 Contact: John Quackenbush
 The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850, USA
 Tel: 301 838 3528
 Fax: 301 838 0208
 Email: johnq@tigr.org
 Plate: 186
 Seq primer: Reverse.

FEATURES
 source
 1. 608
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_lib="MAGE resequences, MACH"
 /note="Vector: pBluescriptSKm"

BASE COUNT 191 a 114 c 117 g 186 t
 ORIGIN
 Query Match 2.1%; Score 20; DB 10; Length 608;
 Best Local Similarity 100.0%; Pred. No. 29;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 891 TCTGAGAAAACCTGGAATA 910
 |||
 Db 111 TCTGAGAAAACCTGGAATA 130

RESULT 10
 AV386359/c 621 bp mRNA linear EST 27-OCT-1999
 LOCUS AV386359 Halocynthia roretzi Petillized egg Halocynthia roretzi
 DEFINITION AV386359 Halocynthia roretzi Petillized egg Halocynthia roretzi
 cDNA clone 009604_3 3', mRNA sequence.
 ACCESSION AV386359

VERSION	AV286359.1	GI:6131416
KEYWORDS	EST	
SOURCE	Halocynthia roretzi.	
ORGANISM	Halocynthia roretzi	
	Eukaryota; Metazoa; Chordata; Urochordata; Ascidiacea;	
	Stolidobranchia; Pyuridae; Halocynthia.	
REFERENCE	1 (bases 1 to 621)	
AUTHORS	Makabe,K.W.	
TITLE	Halocynthia roretzi EST	
JOURNAL	unpublished (1999)	
COMMENT	Contact: Kazuhiro W. Makabe Department of Zoology, Graduate School of Science Kyoto University Sakyo-Ku, Kyoto, Kyoto 606-8502, Japan Tel: 81-75-753-4095 Fax: 81-75-705-1113 Email: kwmakabe@scidjian.zool.kyoto-u.ac.jp.	
FEATURES	Location/Qualifiers	
Source	1..621	
	/organism="Halocynthia roretzi"	
	/db_xref="taxon:7729"	
	/clone_1lb="009G04.3"	
	/clone_1lb="Halocynthia roretzi fertilized egg"	
	/dev_stage="fertilized egg"	
	/note="Organ: embryo"	
BASE COUNT	196 a 99 c 102 g 224 t	
ORIGIN		
Query Match	2.1%; Score 20; DB 10; Length 621;	
Best Local Similarity	100.0%; Pred. No. 29;	
Matches	20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
Oy	384 TCTATTTTTGAATTCGA 403	
Db	465 TCTATTTTGAATTCGA 446	
RESULT 11		
LOCUS	AQ577386 673 bp DNA linear GSS 02-JUN-1999	
DEFINITION	nbx0090P08f CUGI Rice BAC Library Oryza sativa genomic clone	
ACCESSION	AQ577386	
VERSION	AQ577386.1	
KEYWORDS	GSS.	
SOURCE	Oryza sativa.	
ORGANISM	Oryza sativa	
	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;	
	Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;	
	Ehrhartoideae; Oryzeae; Oryza.	
	1 (bases 1 to 673)	
	Wing,R.A. and Dean,R.A.	
	A BAC End Sequencing Framework to Sequence the Rice Genome	
	Unpublished (1998)	
REFERENCE	Contact: Wing RA	
AUTHORS	Clemson University Genomics Institute	
TITLE	Clemson University	
JOURNAL	100 Jordan Hall, Clemson, SC 29634, USA	
COMMENT	Tel: 864 656 7288	
	Fax: 864 656 4293	
	Email: rwing@clemson.edu	
	Seq primer: TAATACGACTCACTATAGG	
	Class: BAC ends	
	High quality sequence stop: 280.	
FEATURES	Location/Qualifiers	
Source	1..673	
	/organism="Oryza sativa"	
	/strain="Japonica"	
	/cultivar="Nipondbare"	
	/db_xref="taxon:4530"	
	/clone_1lb="nbx0090P08f"	
	/clone_1lb="CUGI Rice BAC Library"	
	/tissue_type="Leaf"	

```

/lab host="E. coli DH10B"
/vector="vector: pBelOAC11: Site.1: HindIII: Site.2:
HindIII: Rice is one of two most popular grains in the
world. Half of the world population especially those
inhabiting highly populated areas of the humid tropics
and subtropics, rely on rice as their primary source of
carbohydrate. Monocotyledonous rice is a diploid plant
(2n=24) with a haploid genome equivalent of 431 Mbp
(Arunamathan and Earle, 1991). The relatively small
genome of rice, three times larger than that of
Arabidopsis, makes it suitable for genomic studies. In
order to facilitate positional cloning, physical mapping
and genome sequencing of rice, we have constructed a BAC
library from Oryza sativa, Nipponbare variety. The
library contains 36,864 clones with an average insert size
of 128.5 Kb providing 10.9 haploid genome equivalents. The
deep coverage allows the isolation a particular sequence
with a probability of 99.9 %. Two high density filters,
each containing 18,432 clones (doubly spotted), represent
the whole library for colony screening."

BASE COUNT      231 a      121 c      152 g      169 t

ORIGIN

Query Match      2.1%; Score 20; DB 17; Length 673;
Best Local Similarity 100.0%; Prod. No. 29;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy      15      TTTTGAAGAACAAATAT 34
          |||||
          TTTTGAAGAACAAATAT 617

RESULT 12
BG068343/c      713 bp      mRNA      linear      EST 26-JAN-2001
LOCUS
DEFINITION      H3064D12-3' NIA Mouse 15K cDNA Clone Set Mus musculus cDNA clone
H3064D12 3' mRNA sequence.
BG068343
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Mus musculus.
Mus musculus.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 713)
Kargul,G.J., Dudekula,D.B., Qian,Y., Lam,M.K., Jaradat,S.A., Tanaka
,T.S., Carter,M.G. and Ko,M.S.H.
Verification and initial annotation of NIA mouse 15K cDNA clone set
Unpublished (2001)
Other_ESTs: H3064D12-5
Contact: George J. Kargul
Laboratory of Genetics
National Institute on Aging/National Institutes of Health
333 Cassell Drive, Suite 4000, Baltimore, MD 21224-6620, USA
Email: cdna@jgsun.grc.nia.nih.gov
This clone set has been freely distributed to the community. Please
visit http://jgsun.grc.nia.nih.gov/cDNA/15K.html for details.
Plate: H3064 row: D column: 12
Seq primer: -21M13 Forward
High quality sequence stop: 713
POLYA=yes.
Location/Qualifiers
1..713
/oranism="Mus musculus"
/strain="C57BL/6J"
/db_xref="naest:H3064D12-3"
/db_xref="taxon:10090"
/clone="H3064D12"
/clone_lib="NIA Mouse 15K cDNA Clone Set"
/sex="Clones arrayed from a variety of cDNA libraries"
/dev_stage="Clones arrayed from a variety of cDNA
libraries"
/lab_host="DH10B"

```

/note="Vector: pSPORT1, Site_1: SalI; Site_2: NotI; This clone is among a rearrayed set of 15,247 clones from 11 embryo cDNA libraries (including preimplantation stage embryos from unfertilized egg to blastocyst, embryonic part of E7.5 embryos, extraembryonic part of E7.5 embryos , and E12.5 female mesonephros/gonad) and one newborn ovary cDNA library. Average insert size 1.5 kb. All source libraries are cloned unidirectionally with Oligo(dT) -Not primers. References include: (1) Genome-wide expression profiling of mid-gestation placenta and embryo using a 15,000 mouse developmental cDNA microarray, 2000, Proc. Natl. Acad. Sci. U S A, 97: 9127-9132; (2) Large-scale cDNA analysis reveals phased gene expression patterns during preimplantation mouse development, 2000, Development, 127: 1737-1749; (3) Genome-wide mapping of unselected transcripts from extraembryonic tissue of 7.5-day mouse embryos reveals enrichment in the t-complex and under-representation on the X chromosome, 1998, Hum. Mol. Genet 7: 1967-1978."

BASE COUNT 220 a 112 g 255 t 1 others
ORIGIN

Query Match 2.1%; Score 20; DB 12; Length 713;
Best Local Similarity 100.0%; Pred.No. 29;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 336 TGCCTTTGGAACAATTC 355
|||||
Db 487 TGCCTTTGGAACAATTC 468

RESULT 13
AO914192 769 bp DNA linear GSS 02-DEC-1999
LOCUS nbe004/H17f CUGI Rice BAC Library (ECORI) Oryza sativa genomic
DEFINITION clone nbe004/H17f, DNA sequence.
ACCESSION AO914192
VERSION AO914192.1 GI:6510708
KEYWORDS GSS.
SOURCE Oryza sativa.
ORGANISM Oryza sativa.
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzaceae; Oryza.

REFERENCE 1 (bases 1 to 769)
AUTHORS Wing, R.A. and Dean, R.A.
TITLE A BAC End Sequencing Framework to Sequence the Rice Genome
JOURNAL Unpublished (1998)
COMMENT Contact: Wing RA
Clemson University Genomics Institute
100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 7288
Fax: 864 656 4293
Email: twing@clemson.edu
Seq primer: TATATGACTGACTATAGGG
Class: BAC ends
High quality sequence start: 51
High quality sequence stop: 412.
Location/Qualifiers
1..769
/organism="Oryza sativa"
/strain="Japonica"
/cultivar="Nipponbare"
/db_xref="taxon:4530"
/clone="nbe004/H17f"
/clone_lib="CUGI Rice BAC Library (ECORI)"
/tissue_type="Leaf"
/lab_host="E. coli DH10B"
/note="Vector: PBACindigo; Site_1: EcoRI; Site_2: EcoRI;
Rice is the most important food crop in the world. Half of
the world population, especially those inhabiting highly
populated areas of the humid tropics and subtropics, rely

FEATURES

source

on rice as their primary source of carbohydrate. Monocotyledonous rice is a diploid plant (2n=24) with a haploid genome equivalent of 431 Mbp (Arumuganathan and Earle, 1991). The relatively small genome of rice, three times larger than that of Arabidopsis, makes it suitable for genomic studies. In order to facilitate positional cloning, physical mapping and genome sequencing of rice, we have constructed a BAC library from Oryza sativa, Nipponbare variety using EcoRI as the cloning enzyme. The library contains 55,296 clones with an average insert size of 121 kb providing approximately 15 haploid genome equivalents. The deep coverage allows the isolation a particular sequence with a probability of 99.9 %. Three high density filters, each containing 18,432 clones (doubly spotted), represent the whole library for colony screening and can be requested from the Clemson University BAC/EST Resource Center (www.genome.clemson.edu)."

BASE COUNT 242 a 153 c 161 g 212 t 1 others
ORIGIN

Query Match 2.1%; Score 20; DB 17; Length 769;
Best Local Similarity 100.0%; Pred.No. 29;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 450 TGAGGAGCTCTAAAGTTA 469
|||||
Db 277 TGAGGAGCTCTAAAGTTA 296

RESULT 14
AO871693 771 bp DNA linear GSS 03-NOV-1999
LOCUS nbe0044J03r CUGI Rice BAC Library (ECORI) Oryza sativa genomic
DEFINITION clone nbe0044J03r, DNA sequence.
ACCESSION AO871693
VERSION AO871693.1 GI:6222144
KEYWORDS GSS.
SOURCE Oryza sativa.
ORGANISM Oryza sativa.
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzaceae; Oryza.

REFERENCE 1 (bases 1 to 771)
AUTHORS Wing, R.A. and Dean, R.A.
TITLE A BAC End Sequencing Framework to Sequence the Rice Genome
JOURNAL Unpublished (1998)
COMMENT Contact: Wing RA
Clemson University Genomics Institute
100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 7288
Fax: 864 656 4293
Email: twing@clemson.edu
Seq primer: GGAACAGCTATGACCATG
Class: BAC ends
High quality sequence start: 28
High quality sequence stop: 416.
Location/Qualifiers
1..771
/organism="Oryza sativa"
/strain="Japonica"
/cultivar="Nipponbare"
/db_xref="taxon:4530"
/clone="nbe0044J03r"
/clone_lib="CUGI Rice BAC Library (ECORI)"
/tissue_type="Leaf"
/lab_host="E. coli DH10B"
/note="Vector: PBACindigo; Site_1: EcoRI; Site_2: EcoRI;
Rice is the most important food crop in the world. Half of
the world population, especially those inhabiting highly
populated areas of the humid tropics and subtropics, rely
on rice as their primary source of carbohydrate.
Monocotyledonous rice is a diploid plant (2n=24) with a

FEATURES

source

haploid genome equivalent of 431 Mbp (Arumuganathan and Earle, 1991). The relatively small genome of rice, three times larger than that of Arabidopsis, makes it suitable for genomic studies. In order to facilitate positional cloning, physical mapping and genome sequencing of rice, we have constructed a BAC library from *Oryza sativa*. The Nipponbare variety using EcoRI as the cloning enzyme. The library contains 55,296 clones with an average insert size of 121 kb providing approximately 15 haploid genome equivalents. The deep coverage allows the isolation a particular sequence with a probability of 99.9 %. Three high density filters, each containing 18,432 clones (doubly spotted), represent the whole library for colony screening and can be requested from the Clemson University BAC/EST Resource Center (www.genome.clemson.edu)."

BASE COUNT 240 a 153 c 164 g 212 t 2 others

Query Match 2.1%; Score 20; DB 17; Length 771;
Best Local Similarity 100.0%; Pred. No. 29;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 450 TGAGGAGCTCAAAAGTTA 469
|||||
Db 309 TGAGGAGCTCAAAAGTTA 328

RESULT 15
CNS02ZOR 784 bp DNA linear GSS 15-MAY-2000
LOCUS Tetradon nigroviridis genome survey sequence T7 end of clone
DEFINITION 183K19 of library 6 from Tetradon nigroviridis, genomic survey
sequence.
AL221292.1 GI:7880111
GSS: genome survey sequence.
Tetradon nigroviridis.
Tetradon nigroviridis.
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei;
Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
Tetraodontidae; Tetraodon.
1 (bases 1 to 784)
Roest-Crollius, H., Jallion, O., Dasilva, C., Bouneau, L., Fisher, C.,
Bernot, A., Fizames, C., Wincker, P., Brothier, P., Quetier, F.,
Sautin, W. and Weissenbach, J.
Human gene number estimate provided by genome wide analysis using
Tetradon nigroviridis DNA sequence
Unpublished
2 (bases 1 to 784)
Roest-Crollius, H., Jallion, O., Dasilva, C., Fizames, C., Fisher, C.,
Bouneau, L., Billault, A., Quetier, F., Sautin, W., Bernot, A. and
Weissenbach, J.
Characterization and repeat analysis of the compact genome of the
freshwater pufferfish Tetradon nigroviridis
Unpublished
3 (bases 1 to 784)
Genoscope.
Direct Submission
Submitted (12-APR-2000)
This sequence is a single read and was generated as part of a large
scale clone-end sequencing project of the Tetradon nigroviridis
genome. For more information, please take a look at
<http://www.genoscope.cns.fr/Tetradon>.

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
FEATURES
source
1. 784
/organism="Tetradon nigroviridis"
/db_xref="taxon:99883"
/clone_1ib="G"
/clone_1ib="G"
/note="Genoscope sequence ID : COAG183AF10LPI-end : T7"
BASE COUNT 226 a 148 c 160 g 250 t

Query Match 2.1%; Score 20; DB 17; Length 784;
Best Local Similarity 100.0%; Pred. No. 29;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 549 TGTCCTTGGAAATGGAAAAA 568
|||||
Db 605 TGTCCTTGGAAATGGAAAAA 586

RESULT 16
A2749869 851 bp DNA linear GSS 25-JAN-2001
LOCUS RPCI-24-97H19.TV RPCI-24 Mus musculus genomic clone RPCI-24-97H19,
DNA sequence.
DEFINITION
A2749869
A2749869 GI:12535028
GSS.
house mouse.
Mus musculus
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 851)
Zhao, S., Nierman, W., Malek, J., Shatsman, S., Akirret, B., Levins, M.,
Tsegaye, G., Geer, K., Krol, M., Shwartsbeyn, A., Gebregorjys, E.,
Russell, D., de Jong, P. and Fraser, C.M.
Mouse BAC End Sequences from Library RPCI-24
Unpublished (1999)
Other_GSSs: RPCI-24-97H19.TV
Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhaot@igf.org
Clones are derived from the mouse BAC library RPCI-24. For BAC
library availability, please contact Pieter de Jong
(pdejong@mail.cho.org). Clones may be purchased from BACPAC
Resources (<http://www.choi.org/bacpac/orderingframe.htm>). BAC end
pages: http://www.tlgr.org/ldb/bac_ends/mouse/bac_end_intro.html
Plate: 97 row: H column: 19
Seq primer: SP6
Clas: BAC ends.

FEATURES
source
Location/Qualifiers
1. 851
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="RPCI-24-97H19"
/clone_1ib="RPCI-24"
/sex="Male"
/cell_type="Spleen/Brain"
/note="Vector: pPARBAC1; Site 1: BamHI; Site 2: BamHI;
RPCI-24 Mouse BAC Library produced by Pieter de Jong. The
library was cloned in the pPARBAC1 cloning vector at the
BamHI sites using MboI partially digested male C57BL/6J
DNA."

BASE COUNT 307 a 143 c 103 g 298 t

Query Match 2.1%; Score 20; DB 17; Length 851;
Best Local Similarity 100.0%; Pred. No. 30;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 336 TGCTCTTGGAAACAAATG 355
|||||
Db 799 TGCTCTTGGAAACAAATG 818

RESULT 17
CNS04SPV/c 873 bp DNA linear GSS 24-MAY-2000
LOCUS

DEFINITION Tetraodon nigroviridis genome survey sequence T7 end of clone 007P20 of library H from Tetraodon nigroviridis, genomic survey sequence.

ACCESSION AL305500

VERSION AL305500.1 GI:8198207

KEYWORDS GSS; genome survey sequence.

SOURCE Tetraodon nigroviridis.

ORGANISM Tetraodon nigroviridis

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei; Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes; Tetraodontidae; Tetraodon.

REFERENCE 1 (bases 1 to 873)

AUTHORS Roest-Crollius,H., Jallion,O., Dasilva,C., Bouneau,L., Fisher,C., Bernot,A., Fizames,C., Winkler,P., Brotlier,P., Quetier,F., Sautin,W. and Weissenbach,J.

TITLE Human gene number estimate provided by genome wide analysis using Tetraodon nigroviridis DNA sequence

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 873)

AUTHORS Roest-Crollius,H., Jallion,O., Dasilva,C., Fizames,C., Fisher,C., Bouneau,L., Billault,A., Quetier,F., Sautin,W., Bernot,A. and Weissenbach,J.

TITLE Characterization and repeat analysis of the compact genome of the freshwater pufferfish Tetraodon nigroviridis

JOURNAL Unpublished

REFERENCE 3 (bases 1 to 873)

AUTHORS Direct Submission

TITLE Submitted (12-APR-2000)

COMMENT This sequence is a single read and was generated as part of a large scale clone-end sequencing project of the Tetraodon nigroviridis genome. For more information, please take a look at <http://www.genoscope.cns.fr/Tetraodon.location/Qualifiers>

FEATURES

source

1..873

/organism="Tetraodon nigroviridis"

/db_xref="taxon:99883"

/clone="007P20"

/clone.lib="H"

/note="Genoscope sequence ID : COBH007DH10XD1-end : T7"

BASE COUNT 204 a 232 c 197 g 239 t 1 others

ORIGIN

Query Match 2.1%; Score 20; DB 17; Length 873;

Best Local Similarity 100.0%; Pred. No. 30;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 82 GGAGACTGGGAGTGTGAT 101

|||||

Db 662 GGAGACTGGGAGTGTGAT 643

RESULT 18

AK012307/c 1539 bp mRNA linear HTC 19-JAN-2002

LOCUS AK012307

DEFINITION Mus musculus 11 days embryo whole body cDNA, RIKEN full-length enriched library, clone:2700031606:trans-prenyltransferase, full insert sequence.

ACCESSION AK012307

VERSION AK012307.1 GI:12848964

KEYWORDS HTC; CAP trapper.

SOURCE Mus musculus (strain:C57BL/6J) 11 days embryo cDNA to mRNA, clone.lib:RIKEN full-length enriched mouse cDNA library clone:2700031606.

ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1

AUTHORS Carninci,P. and Hayashizaki,Y.

TITLE High-efficiency full-length cDNA cloning

JOURNAL Meth. Enzymol. 303, 19-44 (1999)

MEDLINE 99279253

PUBMED 10349636

REFERENCE 2

AUTHORS Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.

TITLE Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes

JOURNAL Genome Res. 10 (10), 1617-1630 (2000)

MEDLINE 20493374

PUBMED 11042159

REFERENCE 3

AUTHORS Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P., Konno,H., Akiyama,J., Nishi,K., Kitsuai,T., Tashiro,H., Itoh,M., Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishino,T., Harada,A., Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K., Fujiwara,S., Inoue,K., Togawa,Y., Izawa,M., Ohata,E., Matsubuchi,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.

TITLE RIKEN integrated sequence analysis (RISA) system-384-format sequencing pipeline with 384 multicapillary sequencer

JOURNAL Genome Res. 10 (11), 1757-1771 (2000)

MEDLINE 20530913

PUBMED 11076861

REFERENCE 4

AUTHORS Kawai,J., Shinagawa,A., Shibata,K., Yoshino,M., Itoh,M., Ishii,Y., Arakawa,T., Hara,A., Fukunishi,Y., Konno,H., Adachi,J., Fukuda,S., Aizawa,K., Izawa,M., Nishi,K., Kiyosawa,H., Kondo,S., Yamanaoka,I., Saito,T., Okazaki,Y., Gojobori,T., Bono,H., Kasukawa,T., Saito,R., Kadota,K., Matsuda,H., Ashburner,M., Batalov,S., Casavant,T., Fleischmann,W., Gaasterland,T., Gissi,C., King,B., Kochiya,H., Kuehl,P., Lewis,S., Matsuo,Y., Nikaido,I., Pesole,G., Quackenbush,J., Schriml,L.M., Stabili,F., Suzuki,R., Tomita,M., Wagner,L., Washio,T., Sakai,K., Okido,T., Furuno,M., Aono,H., Baldarelli,R., Barsh,G., Blake,J., Boffelli,D., Bojunga,N., Carninci,P., de Bonaldo,M.F., Brownstein,M.J., Bul,C., Fletcher,C., Fujita,M., Gariboldi,M., Gustincich,S., Hill,D., Hofmann,M., Hume,D.A., Kamiya,M., Lee,N.H., Lyons,P., Marchionni,L., Mashima,J., Mazzarelli,J., Mombauts,P., Nordone,P., Rato,B., Ringwald,M., Rodriguez,I., Sakamoto,N., Sasaki,H., Sato,K., Schonbach,C., Seya,T., Shibata,Y., Storch,K.F., Suzuki,H., Toyooka,K., Wang,K.H., Wetz,C., Whitaker,C., Wilming,L., Wynshaw-Boris,A., Yoshida,K., Hasegawa,Y., Kawai,H., Kohsuki,S. and Hayashizaki,Y.

TITLE Functional annotation of a full-length mouse cDNA collection

JOURNAL Nature 409 (6821), 685-690 (2001)

MEDLINE 21085660

PUBMED 11217851

REFERENCE 5 (bases 1 to 1539)

AUTHORS Adachi,J., Aizawa,K., Akahira,S., Akinura,T., Aono,H., Arai,A., Arakawa,T., Baldarelli,R., Bono,H., Brownstein,M., Bul,C., Carninci,P., Fukuda,S., Fukunishi,Y., Furuno,M., Hanagaki,T., Hara,A., Hayatsu,N., Hill,D., Hiramoto,K., Hiraoaka,T., Horii,F., Hume,D., Imotani,K., Ishii,Y., Itoh,M., Izawa,M., Kasukawa,T., Kato,H., Kawai,J., Kojima,Y., Konno,H., Kouda,M., Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Nishi,K., Nomura,S., Numazaki,R., Ohno,M., Okazaki,Y., Okido,T., Owa,C., Quackenbush,J., Saito,H., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki,D., Schriml,L., Shibata,K., Shibata,Y., Shingawa,A., Shitaki,T., Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F., Tanaka,T., Tejima,Y., Toya,T., Yamamura,T., Yamanaoka,I., Yasunishi,A., Yoshida,K., Yoshino,M., Muramatsu,M. and Hayashizaki,Y.

TITLE Submitted (10-JUN-2000) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Group (GSG), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail:genome-res@gsr.riken.go.jp, URL:<http://genome.gsc.riken.go.jp/>, Tel:81-45-503-9222, Fax:81-45-503-9216)

COMMENT Please visit our web site (<http://genome.gsc.riken.go.jp/>) for further details.

cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken

Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. First strand cDNA was primed with a primer 15' GAGAGAGAGAGATCCAGAGCTCTTTTCTTTTCTTTT 3', cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. Second strand cDNA was prepared with the primer adapter of sequence 15' GAGAGAGAGATCTCGATTAAATTAATATCCCCCCCC 3'. cDNA was cleaved with XhoI and SstI. Cloning sites, 5' end: XhoI; 3' end: SstI. Host: SOLR.

FEATURES
source

Location/Qualifiers
1. .1539
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="FANTOM_DB:2700031G06"
/db_xref="MGD:MGI:1896014"
/db_xref="taxon:10090"
/clone="2700031G06"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="11 days embryo"
257. .880
/note="data source:MGD, source key:MGI:189278, evidence:ISS
putative
trans-prenyltransferase"
/codon_start=1
/protein_id="BAB28153.1"
/db_xref="GI:12848965"
/db_xref="MGD:MGI:1889278"
/translation="MCSGROGSGEKSDPEKLGWRDLKGLYEDIRKELHISTRELKDMSEYFDGKGRKFRPIYVLMARACNIHNNHNRMOASORSIALVAEMHTATLVHDDVIDDASSRRGKHTVKNIMGKRAVLAGDLILSAAYALARTNTAVSMLAQYIEDLV RGEFLQLSKSENEERFAHYLEKTFKASTLIANSCKAVCTFLRSS"
1517. .1522
/note="putative"
1539
/note="putative"

polyA_signal
polyA_site
BASE COUNT 459 a 328 c 317 g 434 t 1 others
ORIGIN

Query Match 2.1%; Score 20; DB 11; Length 1539;
Best local Similarity 100.0%; Pred. No. 31;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 310 CCAGAGAGACACAGGAAAG 329
|||||
DB 1030 CCAGAGAGACACAGGAAAG 1011

Search completed: April 26, 2003, 11:51:59
Job time : 1116 secs

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 26, 2003, 11:23:08 ; Search time 46 Seconds
(without alignments)
6360.209 Million cell updates/sec

Title: US-09-880-729A-1

Perfect score: 954

Sequence: 1 atgggtgtgtatccctttga.....ggagagatagcatgataa 954

Scoring table: OLIGO_NUC

Gapop 60.0 , Gapext 60.0

Searched: 441362 seqs, 153338381 residues

Word size : 20

Total number of hits satisfying chosen parameters: 8

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 1000 summaries

Database :

Issued_Patents_NA:*
1: /cgn2_6/prodata/1/lna/5A.COMB.seq:*
2: /cgn2_6/prodata/1/lna/5B.COMB.seq:*
3: /cgn2_6/prodata/1/lna/6A.COMB.seq:*
4: /cgn2_6/prodata/1/lna/6B.COMB.seq:*
5: /cgn2_6/prodata/1/lna/CTUS.COMB.seq:*
6: /cgn2_6/prodata/1/lna/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	954	100.0	954	2	US-09-066-075-1
2	954	100.0	954	2	US-08-518-615A-1
3	954	100.0	954	3	US-08-951-889-1
4	954	100.0	954	4	US-09-472-857-1
5	21	2.2	51	2	US-09-066-075-3
6	21	2.2	51	2	US-08-518-615A-3
7	21	2.2	51	3	US-08-951-889-3
8	21	2.2	51	4	US-09-472-857-3

ALIGNMENTS

RESULT 1
US-09-066-075-1
; Sequence 1, Application US/09066075
; Patent No. 5925749
; GENERAL INFORMATION:
; APPLICANT: Mathur, E., et al.
; TITLE OF INVENTION: Carboxymethyl Cellulase from Thermotoga Maritima
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: CARELLA, BYRNE, BAIN, GILFILLAN,
; ADDRESSEE: CECCHI, STEWART & OLSTEIN
; STREET: 6 BECKER FARM ROAD
; CITY: ROSELAND
; STATE: NEW JERSEY

COUNTRY: USA
ZIP: 07068
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 INCH DISKETTE
COMPUTER: IBM PS/2
OPERATING SYSTEM: MS-DOS
SOFTWARE: WORD PERFECT 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/066,075
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/518,615
FILING DATE: August 23, 1995
ATTORNEY/AGENT INFORMATION:
NAME: FERRARO, GREGORY D.
REGISTRATION NUMBER: 36,134
REFERENCE/DOCKET NUMBER: 331400-20
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201-994-1700
TELEFAX: 201-994-1744
INFORMATION FOR SEQ. ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 954 BASE PAIRS
TYPE: NUCLEIC ACID
STRANDEDNESS: SINGLE
TOPOLOGY: LINEAR
MOLECULE TYPE: CDNA
US-09-066-075-1

Query Match 100.0%; Score 954; DB 2; Length 954;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 954; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	ATGGGTGTGTATCCCTTTTGAAGCAAAATATGGGAAGGCAATTAATAGCAAAAT	60
DB	1	ATGGGTGTGTATCCCTTTTGAAGCAAAATATGGGAAGGCAATTAATAGCAAAAT	60
QY	61	GGCGTTGAACACCAATGAGGAGCTGGAGTGTGATTAAGATGCTTTCGAC	120
DB	61	GGCGTTGAACACCAATGAGGAGCTGGAGTGTGATTAAGATGCTTTCGAC	120
QY	121	ATTATTAAGAGCCGGTTCTCTCATGTTGCAATTCGAATTAAGATGATAGCAGCT	180
DB	121	ATTATTAAGAGCCGGTTCTCTCATGTTGCAATTCGAATTAAGATGATAGCAGCT	180
QY	181	TACGCGTTCTCTTATTAATCATGATCGCTTCTTCAAAAGAGTGATGAAGTATA	240
DB	181	TACGCGTTCTCTTATTAATCATGATCGCTTCTTCAAAAGAGTGATGAAGTATA	240
QY	241	AAGGAGCCCTGAAAGAGAGAGCGCTGTGCTATAATATTCATCATCGAGGATTA	300
DB	241	AAGGAGCCCTGAAAGAGAGAGCGCTGTGCTATAATATTCATCATCGAGGATTA	300
QY	301	ATGAATGATCCAGAGAACACAGAGAAAGATTTCTTGCTTTGGAACAAATTCGTAT	360
DB	301	ATGAATGATCCAGAGAACACAGAGAAAGATTTCTTGCTTTGGAACAAATTCGTAT	360
QY	361	CGTTATTAAGACTATCCCGAAGCTATATTTTGAATTCGATGAAGCTCAGGAAAT	420
DB	361	CGTTATTAAGACTATCCCGAAGCTATATTTTGAATTCGATGAAGCTCAGGAAAT	420
QY	421	CTTACGCCGGAATAATGGAATGGAAGCTGAGAGGCTTAAGATTAAGATCAAT	480
DB	421	CTTACGCCGGAATAATGGAATGGAAGCTGAGAGGCTTAAGATTAAGATCAAT	480
QY	481	GACAAAAGCACACTAATTAATATAGGACAGTGAATGGGGGTATATGCGCTTGA	540
DB	481	GACAAAAGCACACTAATTAATATAGGACAGTGAATGGGGGTATATGCGCTTGA	540
QY	541	AAACTGTCTGTCACCAAAATGGGAAATTTCTATAGTTACAAATTCACATCAATCCT	600
DB	541	AAACTGTCTGTCACCAAAATGGGAAATTTCTATAGTTACAAATTCACATCAATCCT	600

QY	601	TTGCAATTTACCACATCAAGAGCGTGAATGGGTGGGAACGATCTGGAATGGTTGGGAAG	660
Db	601	TTGCAATTTACCACATCAAGAGCGTGAATGGGTGGGAACGATCTGGAATGGTTGGGAAG	660
QY	661	AAGTGGGATCTCCAGATGATCAGAAACATTTGATTAAGAAATTCATTTTATAGAAGAA	720
Db	661	AAGTGGGATCTCCAGATGATCAGAAACATTTGATTAAGAAATTCATTTTATAGAAGAA	720
QY	721	TGGTCAAAAAAGAACAAAGAACCATTTACATPAGGTGATGTTGGTGGCTTACAGAAAAGCT	780
Db	721	TGGTCAAAAAAGAACAAAGAACCATTTACATPAGGTGATGTTGGTGGCTTACAGAAAAGCT	780
QY	781	GACCTTGAATCAGAAATTAATGACCTCTTTGTGCTGGCGAAATGGAGAAAAGSAGA	840
Db	781	GACCTTGAATCAGAAATTAATGACCTCTTTGTGCTGGCGAAATGGAGAAAAGSAGA	840
QY	841	TGGACCTGGGCATACTGGGAATTTGTCGGGTTTGGTGTATTAAGTACTCTGAGAAA	900
Db	841	TGGACCTGGGCATACTGGGAATTTGTCGGGTTTGGTGTATTAAGTACTCTGAGAAA	900
QY	901	ACCTGGAATTAAGATCTTTTGAAGCTTTAATAGAGAGATATACATTGATAAA	954
Db	901	ACCTGGAATTAAGATCTTTTGAAGCTTTAATAGAGAGATATACATTGATAAA	954

RESULT 2
US-08-518-615A-1
; Sequence 1, Application US/08518615A

```

1  GENERAL INFORMATION:
2  APPLICANT: Mathur, E., et al.
3  TITLE OF INVENTION: Carboxymethyl Cellulase from Thermotoga Maritima
4  NUMBER OF SEQUENCES: 4
5  CORRESPONDENCE ADDRESS:
6  ADDRESSEE: CARELLA, BYRNE, BAIN, GILFILLAN,
7  ADDRESSEE: CECCHI, STEWART & OLSTEIN
8  STREET: 6 BECKER FARM ROAD
9  CITY: ROSELAND
10 STATE: NEW JERSEY
11 COUNTRY: USA
12 ZIP: 07068
13 COMPUTER READABLE FORM:
14 MEDIUM TYPE: 3.5 INCH DISKETTE
15 COMPUTER: IBM PS/2
16 OPERATING SYSTEM: MS-DOS
17 SOFTWARE: WORD PERFECT 5.1
18 CURRENT APPLICATION DATA:
19 APPLICATION NUMBER: US/08/518,615A
20 FILING DATE: August 23, 1995
21 CLASSIFICATION: 435
22 PRIORITY APPLICATION DATA:
23 APPLICATION NUMBER:
24 FILING DATE:
25 ATTORNEY/AGENT INFORMATION:
26 NAME: FERRARO, GREGORY D.
27 REGISTRATION NUMBER: 36,134
28 REFERENCE/DOCKET NUMBER: 331400-20
29 TELECOMMUNICATION INFORMATION:
30 TELEPHONE: 201-994-1700
31 TELEFAX: 201-994-1744
32 INFORMATION FOR SEQ. ID NO. 1:
33 SEQUENCE CHARACTERISTICS:
34 LENGTH: 954 BASE PAIRS
35 TYPE: NUCLEIC ACID
36 STRANDEDNESS: SINGLE
37 TOPOLOGY: LINEAR
38 MOLECULE TYPE: CDNA
39
40 US-08-518-615A-1

```

Query Match	100.0%	Score 954:	DB 2:	length 954:
Best Local Similarity	100.0%	Pred. No. 0:		
Matches 954: Conservative	0:	Mismatches	0:	Gaps 0:

Qy	1	ATGSGTGTGATTCCTTTTGGAAAGCAAAATTTGGGAAAGGCACTTAATTAGGAAT	60
Db	1	ATGGGTGTGTATCTTTTGGAAAGCAAAATTTGGGAAAGGCACTTAATTAGGAAT	60
Qy	61	GGCGTTGAGACCAAAATGAGGGAGACTGGGGAGTGGTATAAAGATGTTCTTCAC	120
Db	61	GGCGTTGAGACCAAAATGAGGGAGACTGGGGAGTGGTATAAAGATGTTCTTCAC	120
Qy	121	ATTATAAAAGACCGGTTTCTCATGTTCCAAATTCCAATTAAGATGAGGTACGCAGT	180
Db	121	ATTATAAAAGACCGGTTTCTCATGTTCCAAATTCCAATTAAGATGAGGTACGCAGT	180
Qy	181	TACGGTTCCTCTTATAAATCATGGATGGCTTCCAAAGAGTGAAGTGAAGTATA	240
Db	181	TACGGTTCCTCTTATAAATCATGGATGGCTTCCAAAGAGTGAAGTGAAGTATA	240
Qy	241	AACGAGCCCTGMAAAGAAGAGACTGGCTGTGCTATAAATATTTCATCATAAGAGATTA	300
Db	241	AACGAGCCCTGMAAAGAAGAGACTGGCTGTGCTATAAATATTTCATCATAAGAGATTA	300
Qy	301	ATGAATGATCCAGAAGACACAGGAAAGATTTCCTGCTTTTGGAAACAAATTGTGAT	360
Db	301	ATGATATGATCCAGAAGACACAGGAAAGATTTCCTGCTTTTGGAAACAAATTGTGAT	360
Qy	361	CGTTTAAAGACTATCCGGAACCTATTTTTTCAAAATTCGATGAACCTCAGGAAAT	420
Db	361	CGTTTAAAGACTATCCGGAACCTATTTTTTCAAAATTCGATGAACCTCAGGAAAT	420
Qy	421	CTTACTCCGGAAAAATGGAATGAACCTGCTTGAGGAGACTCTAAAAAGTTAATAGATCAATT	480
Db	421	CTTACTCCGGAAAAATGGAATGAACCTGCTTGAGGAGACTCTAAAAAGTTAATAGATCAATT	480
Qy	481	GACAAAAGACACATATATTAATAGCACACTGAATGGGGGTATATCGCCCTTGAA	540
Db	481	GACAAAAGACACATATATTAATAGCACACTGAATGGGGGTATATCGCCCTTGAA	540
Qy	541	AAACGTCTGTGCCCCAAAATGGGAAAAAAATCTATAGTACAAATTCCTACTCAATCT	600
Db	541	AAACGTCTGTGCCCCAAAATGGGAAAAAAATCTATAGTACAAATTCCTACTCAATCT	600
Qy	601	TTTCGAATTTACCCATCAAGAGAGCTGAGTGGTGAAGGATCTGAAATGTTGGGAGAGA	660
Db	601	TTTCGAATTTACCCATCAAGAGAGCTGAGTGGTGAAGGATCTGAAATGTTGGGAGAGA	660
Qy	661	AAAGTGGGATCTCCAGATGATCAGAAACATTTGATAGAAGAAATTCATTTTATAGAAGAA	720
Db	661	AAAGTGGGATCTCCAGATGATCAGAAACATTTGATAGAAGAAATTCATTTTATAGAAGAA	720
Qy	721	TGCTCAAAAACAAACAAAGACCAATTTACATPAGGTGAGTTTGGTGCCTACAGAAAAGCT	780
Db	721	TGCTCAAAAACAAACAAAGACCAATTTACATPAGGTGAGTTTGGTGCCTACAGAAAAGCT	780
Qy	781	GACCTTGAATCAAGATATAATGAGACTCTTGTGCTTCGCAAAATGAGAAAAAGAGA	840
Db	781	GACCTTGAATCAAGATATAATGAGACTCTTGTGCTTCGCAAAATGAGAAAAAGAGA	840
Qy	841	TGAGCTGGGCATCTAGGGAATTTTGTCCGGTTTGGTCTTTAATAGACTCTAGAGAAA	900
Db	841	TGAGCTGGGCATCTAGGGAATTTTGTCCGGTTTGGTCTTTAATAGACTCTAGAGAAA	900
Qy	901	ACCTGGAATTAAGATCTTTTACAGCTTTAATAGAGAGATAGACATTTGAATAA	954
Db	901	ACCTGGAATTAAGATCTTTTACAGCTTTAATAGAGAGATAGACATTTGAATAA	954

RESULT 3
 US-08-951-889-1
 ; Sequence 1, Application US/08951889
 ; Patent No. 6008032
 ; GENERAL INFORMATION:
 ; APPLICANT: Mathur, E., et al.
 ; TITLE OF INVENTION: Carboxymethyl Cellulose from

```

: TITLE OF INVENTION: Thermotoga Maritima
: NUMBER OF SEQUENCES: 4
: CORRESPONDENCE ADDRESSES:
: ADDRESSEE: CABELLA, BYRNE, BAIN, GILFILLAN,
: ADDRESSEE: CECCHI, STEWART & OLSTEIN
: STREET: 6 BECKER FARM ROAD
: CITY: ROSELAND
: STATE: NEW JERSEY
: COUNTRY: USA
: ZIP: 07068
: COMPUTER READABLE FORM:
: MEDIUM TYPE: 3.5 INCH DISKETTE
: COMPUTER: IBM PS/2
: OPERATING SYSTEM: MS-DOS
: SOFTWARE: WORD PERFECT 5.1
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/951,889
: FILING DATE:
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 08/518,615
: FILING DATE: August 23, 1995
: ATTORNEY/AGENT INFORMATION:
: NAME: FERRARO, GREGORY D.
: REGISTRATION NUMBER: 36,134
: REFERENCE/DOCKET NUMBER: 331400-20
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 201-994-1700
: TELEFAX: 201-994-1744
: INFORMATION FOR SEQ ID NO: 1:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 954 BASE PAIRS
: TYPE: NUCLEIC ACID
: STRANDEDNESS: SINGLE
: TOPOLOGY: LINEAR
: MOLECULE TYPE: cDNA
: US-08-951-889-1

Query Match      100.0%; Score 954; DB 3; Length 954;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 954; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGGGCTTGATCCTTTGGAAGAACAAATATGGGAAGAGCATTAATATAGAAAT 60
DB 1 ATGGGCTTGATCCTTTGGAAGAACAAATATGGGAAGAGCATTAATATAGAAAT 60
QY 61 GCGCTTGAGGACCAATGAGGAGCTGGGAGTGTATTAAGATGATGCTTCGCAC 120
DB 61 GCGCTTGAGGACCAATGAGGAGCTGGGAGTGTATTAAGATGATGCTTCGCAC 120
QY 121 ATTATAAAGAACCGGTTCTCTCATGTTGCAATTCATAAGATGATGACGCGCT 180
DB 121 ATTATAAAGAACCGGTTCTCTCATGTTGCAATTCATAAGATGATGACGCGCT 180
QY 181 TACGCTTTCCTCTTATAAATCATGATGCTGCTTCTCAAAAGAGCTGATGAATGATA 240
DB 181 TACGCTTTCCTCTTATAAATCATGATGCTGCTTCTCAAAAGAGCTGATGAATGATA 240
QY 241 AAGGAGCCCTGAAAAGAGAGCTGGCTGCTATTAATTCATACATGACGAGGACTTA 300
DB 241 AAGGAGCCCTGAAAAGAGAGCTGGCTGCTATTAATTCATACATGACGAGGACTTA 300
QY 301 ATGAATGATCCAGAAACACAGAGAAAGATTTCTGCTTTGGAACAAATTTGCTGAT 360
DB 301 ATGAATGATCCAGAAACACAGAGAAAGATTTCTGCTTTGGAACAAATTTGCTGAT 360
QY 361 CGTTATAAAGACTATCCCGAAACTCTATTTTGAATTCGATGAATGACCTCAGGAAAT 420
DB 361 CGTTATAAAGACTATCCCGAAACTCTATTTTGAATTCGATGAATGACCTCAGGAAAT 420
QY 421 CTTACTCCGGAATAATGATGAATGATGCTTGAGAGAGCTCTAAAGTTTAAATCAATT 480
DB 421 CTTACTCCGGAATAATGATGAATGATGCTTGAGAGAGCTCTAAAGTTTAAATCAATT 480
```

```

QY 481 GACAAAAGCACATTAATATATAGCCACAGCTGAATGGGGGGGTATATCTGCCCTTGA 540
DB 481 GACAAAAGCACATTAATATATAGCCACAGCTGAATGGGGGGGTATATCTGCCCTTGA 540
QY 541 AAACGTCTGTCCCAAAATGGGAAAAAATTCATAGTATCAATTCACACTACAAATCCT 600
DB 541 AAACGTCTGTCCCAAAATGGGAAAAAATTCATAGTATCAATTCACACTACAAATCCT 600
QY 601 TTGGAATTTACCATCAGAGAGCTGGCTGGAGAGATCTGAGAAATGGTTGGGAAGA 660
DB 601 TTGGAATTTACCATCAGAGAGCTGGCTGGAGAGATCTGAGAAATGGTTGGGAAGA 660
QY 661 AAGTGGGATCTCCAGATGATTCAGAAACATTTGATAGAAATTCATTTATAGAAGA 720
DB 661 AAGTGGGATCTCCAGATGATTCAGAAACATTTGATAGAAATTCATTTATAGAAGA 720
QY 721 TGGTCAAAAAGAACAAAGACCAATTTACATAGTGAGTTGGTGCCTACAGAAAGCT 780
DB 721 TGGTCAAAAAGAACAAAGACCAATTTACATAGTGAGTTGGTGCCTACAGAAAGCT 780
QY 781 GACCTGAATCAGAAATTAATGAGACCTCTGCTGGCGAAATGGAGAAAGAGA 840
DB 781 GACCTGAATCAGAAATTAATGAGACCTCTGCTGGCGAAATGGAGAAAGAGA 840
QY 841 TGGAGCTGGGCATCTGGAATTTGTTCCGTTTGGTGTATGATCTAGAGAAAA 900
DB 841 TGGAGCTGGGCATCTGGAATTTGTTCCGTTTGGTGTATGATCTAGAGAAAA 900
QY 901 ACCTGGAATTAAGATCTTTTGAAGCTTTATATAGAGAGATGACATGTAATA 954
DB 901 ACCTGGAATTAAGATCTTTTGAAGCTTTATATAGAGAGATGACATGTAATA 954
```

RESULT 4

US-09-472-857-1

Sequence 1, Application US/09472857

Patent No. 6245547

GENERAL INFORMATION:

APPLICANT: Mathur, E., et al.

TITLE OF INVENTION: Carboxymethyl Cellulase from

NUMBER OF SEQUENCES: 4

CORRESPONDENCE ADDRESSES:

ADDRESSEE: CABELLA, BYRNE, BAIN, GILFILLAN,

ADDRESSEE: CECCHI, STEWART & OLSTEIN

STREET: 6 BECKER FARM ROAD

CITY: ROSELAND

STATE: NEW JERSEY

COUNTRY: USA

ZIP: 07068

COMPUTER READABLE FORM:

MEDIUM TYPE: 3.5 INCH DISKETTE

COMPUTER: IBM PS/2

OPERATING SYSTEM: MS-DOS

SOFTWARE: WORD PERFECT 5.1

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/472,857

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/951,889

FILING DATE:

APPLICATION NUMBER: 08/518,615

FILING DATE: August 23, 1995

ATTORNEY/AGENT INFORMATION:

NAME: FERRARO, GREGORY D.

REGISTRATION NUMBER: 36,134

REFERENCE/DOCKET NUMBER: 331400-20

TELECOMMUNICATION INFORMATION:

TELEPHONE: 201-994-1700

TELEFAX: 201-994-1744

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:
LENGTH: 954 BASE PAIRS
TYPE: NUCLEIC ACID
STRANDEDNESS: SINGLE
TOPOLOGY: LINEAR
MOLECULE TYPE: CDNA
US-09-472-857-1

Query Match 100.0%; Score 954; DB 4; Length 954;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 954; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 1 ATGGGTGTTGATCCTTTTGAAGAACAATAATTGGGAGAGCATTATATAGCAAT 60
DB 1 ATGGGTGTTGATCCTTTTGAAGAACAATAATTGGGAGAGCATTATATAGCAAT 60
QY 61 GCGGTTGAAGCCCAATAGAGGAGAGCTGGGAGTGTATAAAGATGATCTTTCGAC 120
DB 61 GCGGTTGAAGCCCAATAGAGGAGAGCTGGGAGTGTATAAAGATGATCTTTCGAC 120
QY 121 ATTATAAAGAAAGCCGGTTCTCTCATGTTCCAAATTAAGATGAGTACGACGCT 180
DB 121 ATTATAAAGAAAGCCGGTTCTCTCATGTTCCAAATTAAGATGAGTACGACGCT 180
QY 181 TACGGCTTCCCTTATTAATTCATGATCGCTTCTTCAAAAGAGTGAAGTGAATA 240
DB 181 TACGGCTTCCCTTATTAATTCATGATCGCTTCTTCAAAAGAGTGAAGTGAATA 240
QY 241 AACGAGCCCTGAAAGAGAGAGCTGGCTTGGTATTAATATTCATCATACGAGAGTTA 300
DB 241 AACGAGCCCTGAAAGAGAGAGCTGGCTTGGTATTAATATTCATCATACGAGAGTTA 300
QY 301 ATGAATGATCCAGAAAGACACAAGAAAGATTTCCTCTTTGGAAACAAATTCGTGAT 360
DB 301 ATGAATGATCCAGAAAGACACAAGAAAGATTTCCTCTTTGGAAACAAATTCGTGAT 360
QY 361 CGTTATTAAGATATCCCGAAGCTATTTTGAATTCGATGACCTCAGCAAGAAAT 420
DB 361 CGTTATTAAGATATCCCGAAGCTATTTTGAATTCGATGACCTCAGCAAGAAAT 420
QY 421 CTCTACTCCGGAATAATGGAATGAGTGTGAGAGAGCTCTAAGATTTATAGATCAAT 480
DB 421 CTCTACTCCGGAATAATGGAATGAGTGTGAGAGAGCTCTAAGATTTATAGATCAAT 480
QY 481 GCAAAAACACACATATATTAATGAGCAGAGTGAATGGGGGGGTATTCGCTTGA 540
DB 481 GCAAAAACACACATATATTAATGAGCAGAGTGAATGGGGGGGTATTCGCTTGA 540
QY 541 AAACGTCTGTCCCAAAATGGGAAAAAATTTCTATAGTTACAATTCACACTCAATCCT 600
DB 541 AAACGTCTGTCCCAAAATGGGAAAAAATTTCTATAGTTACAATTCACACTCAATCCT 600
QY 601 TTTCGAATTTACCATATCAAGAGCTGAGTGGTGAAGAGCTGAGAAATGTTGGGAGA 660
DB 601 TTTCGAATTTACCATATCAAGAGCTGAGTGGTGAAGAGCTGAGAAATGTTGGGAGA 660
QY 661 AAGTGGGATCTCCAGATGATCAGAAACATTTGATAGAAATTCATTTTATAGAGAA 720
DB 661 AAGTGGGATCTCCAGATGATCAGAAACATTTGATAGAAATTCATTTTATAGAGAA 720
QY 721 TGGTCAAAAAGAAAGAACCAATTTACATAGTGTGTTGGTCTACAGAAAAGCT 780
DB 721 TGGTCAAAAAGAAAGAACCAATTTACATAGTGTGTTGGTCTACAGAAAAGCT 780
QY 781 GACCTTGAATCAAGATTAATAGGACCTCTTGTGCTTGCGCAATGGAAGAGAGA 840
DB 781 GACCTTGAATCAAGATTAATAGGACCTCTTGTGCTTGCGCAATGGAAGAGAGA 840
QY 841 TGGAGCTGGGATGATGGAATTTTGTTCGGTTTGGTGTATGATCTGAGAAAA 900
DB 841 TGGAGCTGGGATGATGGAATTTTGTTCGGTTTGGTGTATGATCTGAGAAAA 900
QY 901 ACCTGGAATTAAGATCTTTTAGAAGCTTTAATGAGAGAGATGACATTGATAA 954
```

DB 901 ACCTGGAATTAAGATCTTTTAGAAGCTTTAATGAGAGAGATGACATTGATAA 954

RESULT 5

US-09-066-075-3
Sequence 3, Application US/09066075

Patent No. 5925749

GENERAL INFORMATION:

APPLICANT: Mathur, E., et al.

TITLE OF INVENTION: Carboxymethyl Cellulase from Thermotoga Maritima

NUMBER OF SEQUENCES: 4

CORRESPONDENCE ADDRESS:

ADDRESSEE: CARELLA, BYRNE, BAIN, GILFILLAN,

STREET: 6 BECKER FARM ROAD

CITY: ROSELAND

STATE: NEW JERSEY

COUNTRY: USA

ZIP: 07068

COMPUTER READABLE FORM:

MEDIUM TYPE: 3.5 INCH DISKETTE

COMPUTER: IBM PS/2

OPERATING SYSTEM: MS-DOS

SOFTWARE: WORD PERFECT 5.1

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/066,075

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US/08/518,615

FILING DATE: August 23, 1995

ATTORNEY/AGENT INFORMATION:

NAME: FERRARO, GREGORY D.

REGISTRATION NUMBER: 36,134

REFERENCE/DOCKET NUMBER: 331400-20

TELECOMMUNICATION INFORMATION:

TELEPHONE: 201-994-1700

TELEFAX: 201-994-1744

INFORMATION FOR SEQ ID NO: 3:

SEQUENCE CHARACTERISTICS:

LENGTH: 51 BASE PAIRS

TYPE: NUCLEIC ACID

STRANDEDNESS: SINGLE

TOPOLOGY: LINEAR

MOLECULE TYPE: Oligonucleotide

US-09-066-075-3

Query Match 2.2%; Score 21; DB 2; Length 51;

Best Local Similarity 100.0%; Pred. No. 0.86;

Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 1 ATGGGTGTTGATCCTTTTGA 21
DB 31 ATGGGTGTTGATCCTTTTGA 51
```

RESULT 6

US-08-518-615A-3
Sequence 3, Application US/08518615A

Patent No. 5962258

GENERAL INFORMATION:

APPLICANT: Mathur, E., et al.

TITLE OF INVENTION: Carboxymethyl Cellulase from Thermotoga Maritima

NUMBER OF SEQUENCES: 4

CORRESPONDENCE ADDRESS:

ADDRESSEE: CARELLA, BYRNE, BAIN, GILFILLAN,

STREET: 6 BECKER FARM ROAD

CITY: ROSELAND

STATE: NEW JERSEY

COUNTRY: USA

ZIP: 07068

COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 INCH DISKETTE
COMPUTER: IBM PS/2
OPERATING SYSTEM: MS-DOS
SOFTWARE: WORD PERFECT 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/518,615A
FILING DATE: August 23, 1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: FERRARO, GREGORY D.
REGISTRATION NUMBER: 36,134
REFERENCE/DOCKET NUMBER: 331400-20
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201-994-1700
TELEFAX: 201-994-1744
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 51 BASE PAIRS
TYPE: NUCLEIC ACID
STRANDEDNESS: SINGLE
TOPOLOGY: LINEAR
MOLECULE TYPE: Oligonucleotide
US-08-518-615A-3

Query Match 2.2%; Score 21; DB 2; Length 51;
Best Local Similarity 100.0%; Pred. No. 0.86;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ATGGGTGTGATCCTTTGAA 21
|||||
Db 31 ATGGGTGTGATCCTTTGAA 51

RESULT 7
US-08-951-889-3
Sequence 3, Application US/08951889
Patent No. 6008032
GENERAL INFORMATION:
APPLICANT: Mathur, E., et al.
TITLE OF INVENTION: Carboxymethyl Cellulase from
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: CARELLA, BYRNE, BAIN, GILFILLAN,
ADDRESSEE: CECCHI, STEWART & OLSTEIN
STREET: 6 BECKER FARM ROAD
CITY: ROSELAND
STATE: NEW JERSEY
COUNTRY: USA
ZIP: 07068
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 INCH DISKETTE
COMPUTER: IBM PS/2
OPERATING SYSTEM: MS-DOS
SOFTWARE: WORD PERFECT 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/951,889
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/518,615
FILING DATE: August 23, 1995
ATTORNEY/AGENT INFORMATION:
NAME: FERRARO, GREGORY D.
REGISTRATION NUMBER: 36,134
REFERENCE/DOCKET NUMBER: 331400-20
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201-994-1700
TELEFAX: 201-994-1744

INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 51 BASE PAIRS
TYPE: NUCLEIC ACID
STRANDEDNESS: SINGLE
TOPOLOGY: LINEAR
MOLECULE TYPE: Oligonucleotide
US-08-951-889-3

Query Match 2.2%; Score 21; DB 3; Length 51;
Best Local Similarity 100.0%; Pred. No. 0.86;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ATGGGTGTGATCCTTTGAA 21
|||||
Db 31 ATGGGTGTGATCCTTTGAA 51

RESULT 8
US-09-472-857-3
Sequence 3, Application US/09472857
Patent No. 6245547

GENERAL INFORMATION:
APPLICANT: Mathur, E., et al.
TITLE OF INVENTION: Carboxymethyl Cellulase from
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: CARELLA, BYRNE, BAIN, GILFILLAN,
ADDRESSEE: CECCHI, STEWART & OLSTEIN
STREET: 6 BECKER FARM ROAD
CITY: ROSELAND
STATE: NEW JERSEY
COUNTRY: USA
ZIP: 07068
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 INCH DISKETTE
COMPUTER: IBM PS/2
OPERATING SYSTEM: MS-DOS
SOFTWARE: WORD PERFECT 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/472,857
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/951,889
FILING DATE:
APPLICATION NUMBER: 08/518,615
FILING DATE: August 23, 1995
ATTORNEY/AGENT INFORMATION:
NAME: FERRARO, GREGORY D.
REGISTRATION NUMBER: 36,134
REFERENCE/DOCKET NUMBER: 331400-20
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201-994-1700
TELEFAX: 201-994-1744
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 51 BASE PAIRS
TYPE: NUCLEIC ACID
STRANDEDNESS: SINGLE
TOPOLOGY: LINEAR
MOLECULE TYPE: Oligonucleotide
US-09-472-857-3

Query Match 2.2%; Score 21; DB 4; Length 51;
Best Local Similarity 100.0%; Pred. No. 0.86;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ATGGGTGTGATCCTTTGAA 21
|||||
Db 31 ATGGGTGTGATCCTTTGAA 51

Mon Apr 28 09:31:58 2003

us-09-880-729a-1.oli20.rni

Page 6

Search completed: April 26, 2003, 15:27:26
job time : 47 secs

GenCore version 5.1.4-p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 26, 2003, 11:55:53 ; Search time 98 Seconds
(without alignments)
10592.600 Million cell updates/sec

Title: US-09-880-729A-1
Perfect score: 954
Sequence: 1 atgggtgttgatcccttga.....gagagatagcatgataa 954

Scoring table: OLIGO NUC
Gapop 60.0 , Gapext 60.0

Searched: 709820 seqs, 544064369 residues

Word size : 20

Total number of hits satisfying chosen parameters: 6

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Listing first 1000 summaries

Database :

Published Applications_NA.*
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
10: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq:*
11: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
12: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq:*
13: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
14: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	954	100.0	954	US-09-880-729-1	Sequence 1, Appl1
2	21	2.2	51	US-09-880-729-3	Sequence 3, Appl1
3	20	2.1	530	US-09-764-877-824	Sequence 824, Ap
4	20	2.1	1350	US-09-764-877-3648	Sequence 3648, Ap
5	20	2.1	1350	US-09-764-877-3649	Sequence 3649, Ap
6	20	2.1	1350	US-09-764-877-3650	Sequence 3650, Ap

ALIGNMENTS

RESULT 1
US-09-880-729-1
; Sequence 1, Application US/09880729
; Publication No. US20030044956A1
; GENERAL INFORMATION:
; APPLICANT: Mathur, E., et al.
; TITLE OF INVENTION: Carboxymethyl Cellulase from
; Thermotoga Maritima

NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: CABELLA, BYRNE, BAIN, GILFILLAN,
CECHT, STEWART & OLSTEIN
STREET: 6 BECKER FARM ROAD
CITY: ROSELAND
STATE: NEW JERSEY
COUNTRY: USA
ZIP: 07068
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 INCH DISKETTE
COMPUTER: IBM PS/2
OPERATING SYSTEM: MS-DOS
SOFTWARE: WORD PERFECT 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/880,729
FILING DATE: 12-Jun-2001
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/09/472,857
FILING DATE: <Unknown>
APPLICATION NUMBER: 08/951,889
FILING DATE: <Unknown>
APPLICATION NUMBER: 08/518,615
FILING DATE: August 23, 1995
ATTORNEY/AGENT INFORMATION:
NAME: FERRARO, GREGORY D.
REGISTRATION NUMBER: 36,134
REFERENCE/DOCKET NUMBER: 331400-20
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201-994-1700
TELEFAX: 201-994-1744
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 954 BASE PAIRS
TYPE: NUCLEIC ACID
STRANDEDNESS: SINGLE
TOPOLOGY: LINEAR
MOLECULE TYPE: cDNA
SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-880-729-1
Query Match 100.0%; Score 954; DB 9; Length 954;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 954; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
1 ATGGGTGTTGATCCTTTGAAAGAACAAATATTGGGAAGAGCATTAATATAGCAAT 60
1 ATGGGTGTTGATCCTTTGAAAGAACAAATATTGGGAAGAGCATTAATATAGCAAT 60
61 GCGCTTGAACACCAATGAGGAGAGCTGGGAGCTGTGATTAAGATGAGTCTTGAC 120
61 GCGCTTGAACACCAATGAGGAGAGCTGGGAGCTGTGATTAAGATGAGTCTTGAC 120
121 ATTATATAAAGACCCGTTCTCTCATGTTGCAATTCATTAAGATGAGTACGACGT 180
121 ATTATATAAAGACCCGTTCTCTCATGTTGCAATTCATTAAGATGAGTACGACGT 180
181 TACGCGTTCTCTCTTATTAATATGATGAGTCTTCTTCAAAAGAGTGATGAAGTGATA 240
181 TACGCGTTCTCTCTTATTAATATGATGAGTCTTCTTCAAAAGAGTGATGAAGTGATA 240
241 AACGAGACCCCTGAAGAGAGAGCTGGCTGTATTAATATGATGATGAGGAGATTA 300
241 AACGAGACCCCTGAAGAGAGAGCTGGCTGTATTAATATGATGATGAGGAGATTA 300
301 ATGAATGATCCAGAGAACCAAGAGAAATTTCTGCTTTGGAAGCAATTTGCTGAT 360
301 ATGAATGATCCAGAGAACCAAGAGAAATTTCTGCTTTGGAAGCAATTTGCTGAT 360
361 CGTTATTAAGACTATCCCGAACTATTTTGAATTTGGAATTTGGAATGAGCTCAGCAAT 420
361 CGTTATTAAGACTATCCCGAACTATTTTGAATTTGGAATTTGGAATGAGCTCAGCAAT 420

QY	421	CTCTCTCGGAAAATGSAATGACACCTGCTGAGAACCTCTTAAAGATTATTAATCATTAAT	480
Db	421	CTTCTCTCGGAAAATGSAATGACACCTGCTGAGAACCTCTTAAAGATTATTAATCATTAAT	480
QY	481	GACAAAAAGCAGACTATATATATATAGGCACAGCTGATATGSGGGGGGTATATCTGCCCTTGA	540
Db	481	GACAAAAAGCAGACTATATATATATAGGCACAGCTGATATGSGGGGGGTATATCTGCCCTTGA	540
QY	541	AAACGTCGTGCCAAAATGGGAAAAAAATTTCTATCTTCAATTCACCTACTCAATTCCT	600
Db	541	AAACGTCGTGCCAAAATGGGAAAAAAATTTCTATCTTCAATTCACCTACTCAATTCCT	600
QY	601	TTTCGAATTTACCATCAAGAGAGCTAGTGGGTGGAAAGATCTGAGAAATGTTGGGAGAG	660
Db	601	TTTCGAATTTACCATCAAGAGAGCTAGTGGGTGGAAAGATCTGAGAAATGTTGGGAGAG	660
QY	661	AAGTGGGATCTCCAGATGATCATCAAAACATTTGATATAGAAGAAATTCATTTTATAGAAGAA	720
Db	661	AAGTGGGATCTCCAGATGATCATCAAAACATTTGATATAGAAGAAATTCATTTTATAGAAGAA	720
QY	721	TGGTCAAAAAAAGAACAAATAGACCAATTTACATPAGTGGATGATTTGGTGGCTACAGAAAAAGCT	780
Db	721	TGGTCAAAAAAAGAACAAATAGACCAATTTACATPAGTGGATGATTTGGTGGCTACAGAAAAAGCT	780
QY	781	GACCTTGAATCAAGATATAAATGGAACCTCTTTGTGCTGTGCCGAAATGGGAAAAAGAGAG	840
Db	781	GACCTTGAATCAAGATATAAATGGAACCTCTTTGTGCTGTGCCGAAATGGGAAAAAGAGAG	840
QY	841	TGGACCTGGGCATCTAGTGGGAATTTTGTGTCGCGATTTGGTGTGTTATGATACTCTGAGAAAA	900
Db	841	TGGACCTGGGCATCTAGTGGGAATTTTGTGTCGCGATTTGGTGTGTTATGATACTCTGAGAAAA	900
QY	901	ACCTCGAATTAAGATCTTTTGAAGCTTAAATAGAGAGATATACCATTTGAATTA	954
Db	901	ACCTCGAATTAAGATCTTTTGAAGCTTAAATAGAGAGATATACCATTTGAATTA	954

RESULT 2
 US-09-880-729-3
 Sequence 3, Application US/09880729
 Publication No. US20030044955A1
 GENERAL INFORMATION:
 APPLICANT: Mathur, E., et al.
 TITLE OF INVENTION: Carboxymethyl Cellulase from
 Thermotoga Maritima
 NUMBER OF SEQUENCES: 4
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: CECCHI, STEWART & OLSTEIN
 STREET: 6 BECKER FARM ROAD
 CITY: ROSELAND
 STATE: NEW JERSEY
 COUNTRY: USA
 ZIP: 07068
 COMPUTER READABLE FORM:
 MEDIUM TYPE: 3.5 INCH DISKETTE
 COMPUTER: IBM PS/2
 OPERATING SYSTEM: MS-DOS
 SOFTWARE: WORD PERFECT 5.1
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/09/880,729
 FILING DATE: 12-Jun-2001
 CLASSIFICATION: <unknown>
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US/09/472,857
 FILING DATE: <unknown>
 APPLICATION NUMBER: 08/951,889
 FILING DATE: <unknown>
 APPLICATION NUMBER: 08/518,615
 FILING DATE: August 23, 1995
 ATTORNEY/AGENT INFORMATION:
 NAME: FERRARO, GREGORY D.

```

?      REGISTRATION NUMBER: 36,134
?      REFERENCE/DOCKET NUMBER: 331400-20
?      TELECOMMUNICATION INFORMATION:
?      TELEPHONE: 201-994-1700
?      TELEFAX: 201-994-1744
?      INFORMATION FOR SEQ ID NO: 3:
?      SEQUENCE CHARACTERISTICS:
?          LENGTH: 51 BASE PAIRS
?          TYPE: NUCLEIC ACID
?          STRANDEDNESS: SINGLE
?          TOPOLOGY: LINEAR
?      MOLECULE TYPE: Oligonucleotide
?      SEQUENCE DESCRIPTION: SEQ ID NO: 3:
US-09-880-729-3

Query Match          2.2%  Score 21;  DB 9;  Length 51;
Best Local Similarity 100.0%; Pred. NO. 0.62;
Matches 21;  Conservative 0;  Mismatches 0;  Indels 0;  Gaps 0;

Oy      1  ATGGGTGTTGATCCTTTTGA  21
        |||||
Db      31  ATGGGTGTTGATCCTTTTGA  51

```

```

RESULT 3
US-09-764-877-824
; Sequence 824, Application US/09764877
; Patent No. US20020147140A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC005
; CURRENT APPLICATION NUMBER: US/09/764,877
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - refer to PALM or file wrapper
; NUMBER OF SEQ. ID NOS: 4031
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 824
; LENGTH: 530
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-877-824

Query Match
Best Local Similarity 2.1%; Score 20; DB 10; Length 530;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0

QY 891 TCTGAGAAAACCTGGAATA 910
      |||||||
DB 63 TCTGAGAAAACCTGGAATA 82

RESULT 4
US-09-764-877-3648
; Sequence 3648, Application US/09764877
; Patent No. US20020147140A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC005
; CURRENT APPLICATION NUMBER: US/09/764,877
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - refer to PALM or file wrapper
; NUMBER OF SEQ. ID NOS: 4031
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 3648
; LENGTH: 1350
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-877-3648

Query Match
Best Local Similarity 2.1%; Score 20; DB 10; Length 1350;
Matches 100.0%; Pred. No. 2.6; Indels 0; Gaps 0

```

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 891 TCTGAGAAAAAACCCTGGAATA 910
 |||
 Db 866 TCTGAGAAAAAACCCTGGAATA 885

RESULT 5

US-09-764-877-3649
 ; Sequence 3649, Application US/09764877
 ; Patent No. US20020147140A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Rosen et al.
 ; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
 ; FILE REFERENCE: PC005
 ; CURRENT APPLICATION NUMBER: US/09/764,877
 ; PRIORITY FILING DATE: 2001-01-17
 ; Prior application data removed - refer to PALM or file wrapper
 ; NUMBER OF SEQ ID NOS: 4031
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 3649
 ; LENGTH: 1350
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 US-09-764-877-3649

Query Match

2.1%; Score 20; DB 10; Length 1350;
 Best Local Similarity 100.0%; Pred. No. 2.8;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 891 TCTGAGAAAAAACCCTGGAATA 910
 |||
 Db 866 TCTGAGAAAAAACCCTGGAATA 885

RESULT 6

US-09-764-877-3650
 ; Sequence 3650, Application US/09764877
 ; Patent No. US20020147140A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Rosen et al.
 ; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
 ; FILE REFERENCE: PC005
 ; CURRENT APPLICATION NUMBER: US/09/764,877
 ; PRIORITY FILING DATE: 2001-01-17
 ; Prior application data removed - refer to PALM or file wrapper
 ; NUMBER OF SEQ ID NOS: 4031
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 3650
 ; LENGTH: 1350
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 US-09-764-877-3650

Query Match

2.1%; Score 20; DB 10; Length 1350;
 Best Local Similarity 100.0%; Pred. No. 2.8;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 891 TCTGAGAAAAAACCCTGGAATA 910
 |||
 Db 866 TCTGAGAAAAAACCCTGGAATA 885

Search completed: April 26, 2003, 16:05:07
 Job time : 100 secs

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 26, 2003, 10:49:02 ; Search time 1910 Seconds

(without alignments)
14536.166 Million cell updates/sec

Title: US-09-880-729a-1

Perfect score: 954
Sequence: 1 atgggtgtgtatccctttga.....gagagatagcatgataa 954

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 2054640 seqs, 14551402878 residues

Word size : 20

Total number of hits satisfying chosen parameters: 87

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database : GenEmbl:**

1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_pro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vl:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vl:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pin:*
35: em_htg_rtd:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	954	100.0	954	6	AR077712	AR077712 Sequence
2	954	100.0	954	6	AR096552	AR096552 Sequence
3	954	100.0	954	6	AR157661	AR157661 Sequence
4	903	94.7	10244	1	AE001813	AE001813 Thermotog
5	23	2.4	150882	2	AC079067	AC079067 Homo sapi
6	23	2.4	160529	9	AC092659	AC092659 Homo sapi
7	23	2.3	178997	2	AC126611	AC126611 Mus muscu
8	22	2.3	194378	2	AC108661	AC108661 Rattus no
9	21	2.2	51	6	AR077713	AR077713 Sequence
10	21	2.2	51	6	AR096553	AR096553 Sequence
11	21	2.2	51	6	AR157662	AR157662 Sequence
12	21	2.2	106571	9	HS86E14	299572 Human DNA s
13	21	2.2	134757	9	AC007870	AC007870 Genomic s
14	21	2.2	134781	10	AL355177	AL355177 Mouse DNA
15	21	2.2	137154	9	AL513082	AL513082 Human DNA
16	21	2.2	140795	2	AC079562	AC079562 Mus muscu
17	21	2.2	146385	9	AC010534	AC010534 Homo sapi
18	21	2.2	166496	9	AC124046	AC124046 Homo sapi
19	21	2.2	175518	9	AC018846	AC018846 Homo sapi
20	21	2.2	187038	2	AC021179	AC021179 Homo sapi
21	21	2.2	196216	2	AC019228	AC019228 Homo sapi
22	21	2.2	216988	10	AC084217	AC084217 Mus Muscu
23	21	2.2	243381	2	AC125205	AC125205 Mus Muscu
24	20	2.1	849	6	E06193	E06193 CDNA encodl
25	20	2.1	849	6	E06204	E06204 CDNA encodl
26	20	2.1	849	6	E06205	E06205 CDNA encodl
27	20	2.1	849	6	E06206	E06206 CDNA encodl
28	20	2.1	849	6	E06389	E06389 CDNA encodl
29	20	2.1	849	6	E06400	E06400 CDNA encodl
30	20	2.1	849	6	E06401	E06401 CDNA encodl
31	20	2.1	849	6	E06402	E06402 CDNA encodl
32	20	2.1	1280	6	E06200	E06200 CDNA encodl
33	20	2.1	1280	6	E06216	E06216 CDNA encodl
34	20	2.1	1280	6	E06217	E06217 CDNA encodl
35	20	2.1	1280	6	E06396	E06396 CDNA encodl
36	20	2.1	1280	6	E06412	E06412 CDNA encodl
37	20	2.1	1280	6	E06413	E06413 CDNA encodl
38	20	2.1	3564	6	E06203	E06203 CDNA encodl
39	20	2.1	3564	6	E06223	E06223 CDNA encodl
40	20	2.1	3564	6	E06399	E06399 CDNA encodl
41	20	2.1	3564	6	E06419	E06419 CDNA encodl
42	20	2.1	9359	14	AF313916	AF313916 Hepatitis C
43	20	2.1	9431	14	HCCHCPO	D45172 Hepatitis C
44	20	2.1	9431	14	HCPCP	D30613 Hepatitis C
45	20	2.1	9471	6	E06261	E06261 CDNA encodl
46	20	2.1	9471	6	E06457	E06457 CDNA encodl
47	20	2.1	9504	14	AB049099	AB049099 Hepatitis
48	20	2.1	10435	1	AE008634	AE008634 Rickettsi
49	20	2.1	13610	1	AE013181	AE013181 Thermopana
50	20	2.1	39726	3	CEB0413	Z92824 Caenorhabdi
51	20	2.1	46206	9	AL591062	AL591062 Human DNA
52	20	2.1	55577	2	AC014109	AC014109 Drosophila
53	20	2.1	80025	9	AL391355	AL391355 Human DNA
54	20	2.1	84281	2	AC097840	AC097840 Rattus no
55	20	2.1	104753	2	AC114062	AC114062 Rattus no
56	20	2.1	109922	2	AC011671	AC011671 Homo sapi
57	20	2.1	110000	2	CEY3986_2	Continuation (3 of
58	20	2.1	117000	9	AC069222	AC069222 Homo sapi
59	20	2.1	126547	8	AC006586	AC006586 Arabidops
60	20	2.1	126547	2	AC103198	AC103198 Rattus no
61	20	2.1	139571	2	AC095268	AC095268 Rattus no
62	20	2.1	144564	8	AC119071	AC119071 Oryza sat
63	20	2.1	151281	9	AC103660	AC103660 Homo sapi
64	20	2.1	152233	2	AF268619	AF268619 Homo sapi
65	20	2.1	155204	2	AC007926	AC007926 Trypanoso

Pred. No. is the number of results predicted by chance to have a

```

c 66 2.1 155761 2 AP005578
c 67 2.1 158774 2 AP005587
c 68 2.1 159020 2 AC092014
c 69 2.1 159356 2 AL845427
c 70 2.1 162458 2 AC128759
c 71 2.1 166991 2 AC015952
c 72 2.1 166992 2 AC022883
c 73 2.1 168366 9 AC010002
c 74 2.1 169921 9 AC093497
c 75 2.1 170071 9 AC092022
c 76 2.1 170141 9 AC091519
c 77 2.1 170423 2 AC092787
c 78 2.1 171044 2 AC095222
c 79 2.1 174974 2 AC119585
c 80 2.1 179678 2 AC123247
c 81 2.1 198043 2 AC123247
c 82 2.1 217766 10 AC087558
c 83 2.1 222368 2 AC103652
c 84 2.1 246237 3 CER3986A
c 85 2.1 264211 2 AC123867
c 86 2.1 274349 2 AC093623
c 87 2.1 295377 3 AE003532
```

ALIGNMENTS

```

RESULT 1
LOCUS AR077712 954 bp DNA linear PAT 31-AUG-2000
DEFINITION Sequence 1 from patent US 5962258.
ACCESSION AR077712.1 GI:10004458
VERSION
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 954)
AUTHORS Mathur,E.J. and Lam,D.E.
TITLES Carboxymethyl cellulase from thermotoga maritima
JOURNAL Patent: US 5962258-A 1 05-OCT-1993;
FEATURES
source
BASE COUNT 335 a 145 c 221 g 253 t
ORIGIN
```

```

Query Match 100.0%; Score 954; DB 6; Length 954;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 954; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```

QY 1 ATGGGCTTGATCGCTTTGAAAGAACAAATATTGGAGAGGCAATTAATAGAAAT 60
DB 1 ATGGGCTTGATCGCTTTGAAAGAACAAATATTGGAGAGGCAATTAATAGAAAT 60
QY 61 GCGCTTGAAACACCAATGAGGAGACTGGGAGCTGTGATAAAAAGATGAGTCTTGAC 120
DB 61 GCGCTTGAAACACCAATGAGGAGACTGGGAGCTGTGATAAAAAGATGAGTCTTGAC 120
QY 121 ATTATTAAGAAGCGGTTTCTCTCATGTTCGAATTCATTAAGATGAGAGCAGCT 180
DB 121 ATTATTAAGAAGCGGTTTCTCTCATGTTCGAATTCATTAAGATGAGAGCAGCT 180
QY 181 TAGCGCTTGCTCTTAAATATCATGATCGCTTCAAAAGAGTGATGATGATA 240
DB 181 TAGCGCTTGCTCTTAAATATCATGATCGCTTCAAAAGAGTGATGATGATA 240
QY 241 AAGGAGCCCTGANAAGAGAGCTGGCTTGCTATAATATTCATCACTACGAGAGTTA 300
DB 241 AAGGAGCCCTGANAAGAGAGCTGGCTTGCTATAATATTCATCACTACGAGAGTTA 300
QY 301 ATGATATGTCAGAGAAGACACAGAGAAAGATTTCTGCTCTTTGGAAACAATTCGCAT 360
DB 301 ATGATATGTCAGAGAAGACACAGAGAAAGATTTCTGCTCTTTGGAAACAATTCGCAT 360
```

```

DB 301 ATGATATGTCAGAGAAGACACAGAGAAAGATTTCTGCTCTTTGGAAACAATTCGCAT 360
QY 361 CGTTATTAAGACTATCCCGAAGCTCTATTTTGAATTCGATGAGACCTCAGGAAT 420
DB 361 CGTTATTAAGACTATCCCGAAGCTCTATTTTGAATTCGATGAGACCTCAGGAAT 420
QY 421 CTTACTCCGGAATAATGATGATGACTGCTGAGAGAGCTTAAAGTTATAGATCAAT 480
DB 421 CTTACTCCGGAATAATGATGATGACTGCTGAGAGAGCTTAAAGTTATAGATCAAT 480
QY 481 GACAAAGACACATATATATATAGGACACAGCTGAATGGGGGCTATATCTGCCCTTGA 540
DB 481 GACAAAGACACATATATATATAGGACACAGCTGAATGGGGGCTATATCTGCCCTTGA 540
QY 541 AAGCTGCTGCTCCCAATTTGGGAAAAAATCTATAGTTACATTCCTACTACATCTCT 600
DB 541 AAGCTGCTGCTCCCAATTTGGGAAAAAATCTATAGTTACATTCCTACTACATCTCT 600
QY 601 TTGCAATTTACCATCAAGAGAGCTGAGTGGGTGAAGGANTCTGAGAAATGGTTGGAGA 660
DB 601 TTGCAATTTACCATCAAGAGAGCTGAGTGGGTGAAGGANTCTGAGAAATGGTTGGAGA 660
QY 661 AAGTGGGATCTCCAGATGATGAGAAACATTTGATGAGAAATTCATTTATGAGAA 720
DB 661 AAGTGGGATCTCCAGATGATGAGAAACATTTGATGAGAAATTCATTTATGAGAA 720
QY 721 TGCTCAAAAAGACAAAGACCAATTTACATAGCTGAGTTGCTCTACAGAAAGCT 780
DB 721 TGCTCAAAAAGACAAAGACCAATTTACATAGCTGAGTTGCTCTACAGAAAGCT 780
QY 781 GACCTTGAATCAAGATTAATGAGACCTCTTTCGCTTCGCGAATGAGAAAAAGAGA 840
DB 781 GACCTTGAATCAAGATTAATGAGACCTCTTTCGCTTCGCGAATGAGAAAAAGAGA 840
QY 841 TGGAGCTGGGCACTACTGGAATTTTTCGCTTTCGCTTTCGCTTTCGCTTTCGCTTTC 900
DB 841 TGGAGCTGGGCACTACTGGAATTTTTCGCTTTCGCTTTCGCTTTCGCTTTCGCTTTC 900
QY 901 ACCTGGAATTAAGATCTTTTGAAGCTTTTATAGAGAGATGATGATGATGATGATGAT 954
DB 901 ACCTGGAATTAAGATCTTTTGAAGCTTTTATAGAGAGATGATGATGATGATGATGAT 954
```

```

RESULT 2
LOCUS AR096552 954 bp DNA linear PAT 08-SEP-2000
DEFINITION Sequence 1 from patent US 6008032.
ACCESSION AR096552
VERSION AR096552.1 GI:10025457
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 954)
AUTHORS Mathur,E.J. and Lam,D.E.
TITLES Carboxymethyl cellulase from Thermotoga maritima
JOURNAL Patent: US 6008032-A 1 28-DEC-1999;
FEATURES
source
BASE COUNT 335 a 145 c 221 g 253 t
ORIGIN
```

```

Query Match 100.0%; Score 954; DB 6; Length 954;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 954; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```

QY 1 ATGGGCTTGATCGCTTTGAAAGAACAAATATTGGAGAGGCAATTAATAGAAAT 60
DB 1 ATGGGCTTGATCGCTTTGAAAGAACAAATATTGGAGAGGCAATTAATAGAAAT 60
QY 61 GCGCTTGAAACACCAATGAGGAGACTGGGAGCTGTGATAAAAAGATGAGTCTTGAC 120
DB 61 GCGCTTGAAACACCAATGAGGAGACTGGGAGCTGTGATAAAAAGATGAGTCTTGAC 120
```

Db 61 GCGCTTGAGACCAAAATGAGGAGACTGGGAGTGTGATAAAAGATGCTTCTGCAC 120
QY 121 ATTATAAAGAACCGGTTTCTCTCATGTTCGATTCGAATTCGAATAGATGAGATAGCAGCGCT 180
Db 121 ATTATAAAGAACCGGTTTCTCTCATGTTCGATTCGAATTCGAATAGATGAGATAGCAGCGCT 180
QY 181 TAGCGCTTCTCTCTATATAATCATGATGCTTCTTCAAAAAGAGTGAATGATGATA 240
Db 181 TAGCGCTTCTCTCTATATAATCATGATGCTTCTTCAAAAAGAGTGAATGATGATA 240
QY 241 AAGGAGCCCTGAAAAAGAGACTGGCTGTGCTATATAATTCATCATCTACGAGAGTTA 300
Db 241 AAGGAGCCCTGAAAAAGAGACTGGCTGTGCTATATAATTCATCATCTACGAGAGTTA 300
QY 301 ATGAATGATCAGAAAGACCAAGAAAGATTCCTGCTTTTGGAACAATTCGAT 360
Db 301 ATGAATGATCAGAAAGACCAAGAAAGATTCCTGCTTTTGGAACAATTCGAT 360
QY 361 CGTTATTAAGACTATCCCGAACTCTATTTTGAATTCGATGAACTCAGCGAAT 420
Db 361 CGTTATTAAGACTATCCCGAACTCTATTTTGAATTCGATGAACTCAGCGAAT 420
QY 421 CTATCTCCGAAAAATGAGATGAGCTGTGAGAGAGCTTAAAGTATTAAGATCAATT 480
Db 421 CTATCTCCGAAAAATGAGATGAGCTGTGAGAGAGCTTAAAGTATTAAGATCAATT 480
QY 481 GACAAAAAGACACTATTAATTAAGACAGCTGAATGGGGGCTATATCTGCCCTTGA 540
Db 481 GACAAAAAGACACTATTAATTAAGACAGCTGAATGGGGGCTATATCTGCCCTTGA 540
QY 541 AAACGTCTGTCCCAAAATGGGAAAAATTCATAGTTACAATTTCATCTACATCTCT 600
Db 541 AAACGTCTGTCCCAAAATGGGAAAAATTCATAGTTACAATTTCATCTACATCTCT 600
QY 601 TTGGAATTTACCATCAGAGAGCTGAGTGGTGAAGAGATCTGAGAAATGGTGGAGAA 660
Db 601 TTGGAATTTACCATCAGAGAGCTGAGTGGTGAAGAGATCTGAGAAATGGTGGAGAA 660
QY 661 AAGTGGGATCTCCAGATGATCAGAAACATTTGATAGAAAGATTCATTTATAGAGAA 720
Db 661 AAGTGGGATCTCCAGATGATCAGAAACATTTGATAGAAAGATTCATTTATAGAGAA 720
QY 721 TGGTCAAAAAAGACAAAGACCAATTTACATAGTGTGAGTTGGTGCCTACAGAAAA 780
Db 721 TGGTCAAAAAAGACAAAGACCAATTTACATAGTGTGAGTTGGTGCCTACAGAAAA 780
QY 781 GACCTGATCAAGATTAATATGAGTGTGAGTTGGTGCCTACAGAAAA 840
Db 781 GACCTGATCAAGATTAATATGAGTGTGAGTTGGTGCCTACAGAAAA 840
QY 841 TGGAGCTGGGCACTACGGAATTTGTTCGGTTTGGTGTATGATCTGAGAAAA 900
Db 841 TGGAGCTGGGCACTACGGAATTTGTTCGGTTTGGTGTATGATCTGAGAAAA 900
QY 901 ACCTGGAATTAAGATCTTTTGAAGCTTTTAATAGAGAGATGATGATTA 954
Db 901 ACCTGGAATTAAGATCTTTTGAAGCTTTTAATAGAGAGATGATGATTA 954
RESULT 3
AR157661 954 bp DNA linear PAT 17-OCT-2001
LOCUS AR157661
DEFINITION Sequence 1 from patent US 6245547.
ACCESSION AR157661
VERSION AR157661.1 GI:16218624
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 954)
AUTHORS Mathur E.J. and Lam D.E.
TITLE Carboxymethyl cellulose from thermotoga maritima
JOURNAL Patent: US 6245547-A I 12-JUN-2001;

FEATURES
source Location/Qualifiers
BASE COUNT 335 a 145 c 221 g 253 t
ORIGIN
Query Match 100.0%; Score 954; DB 6; Length 954;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 954; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ATGGGTGTTATCTTTTGAAGAACAAAATATGGGAAGAGAGCATTAATATAGAAAT 60
Db 1 ATGGGTGTTATCTTTTGAAGAACAAAATATGGGAAGAGAGCATTAATATAGAAAT 60
QY 61 GCGCTTACACCAAAATGAGGAGAGTGGGAGTGGTGAATTAAGATGATGCTTCTGCAC 120
Db 61 GCGCTTACACCAAAATGAGGAGAGTGGGAGTGGTGAATTAAGATGATGCTTCTGCAC 120
QY 121 ATTATAAAGAACCGGTTTCTCTCATGTTCGATTCGAATTCGAATAGATGAGATAGCAGCGCT 180
Db 121 ATTATAAAGAACCGGTTTCTCTCATGTTCGATTCGAATTCGAATAGATGAGATAGCAGCGCT 180
QY 181 TAGCGCTTCTCTCTATATAATCATGATGCTTCTTCAAAAAGAGTGAATGATGATA 240
Db 181 TAGCGCTTCTCTCTATATAATCATGATGCTTCTTCAAAAAGAGTGAATGATGATA 240
QY 241 AAGGAGCCCTGAAAAAGAGACTGGCTGTGCTATATAATTCATCATCTACGAGAGTTA 300
Db 241 AAGGAGCCCTGAAAAAGAGACTGGCTGTGCTATATAATTCATCATCTACGAGAGTTA 300
QY 301 ATGAATGATCAGAAAGACCAAGAAAGATTCCTGCTTTTGGAACAATTCGAT 360
Db 301 ATGAATGATCAGAAAGACCAAGAAAGATTCCTGCTTTTGGAACAATTCGAT 360
QY 361 CGTTATTAAGACTATCCCGAACTCTATTTTGAATTCGATGAACTCAGCGAAT 420
Db 361 CGTTATTAAGACTATCCCGAACTCTATTTTGAATTCGATGAACTCAGCGAAT 420
QY 421 CTATCTCCGAAAAATGAGATGAGCTGTGAGAGAGCTTAAAGTATTAAGATCAATT 480
Db 421 CTATCTCCGAAAAATGAGATGAGCTGTGAGAGAGCTTAAAGTATTAAGATCAATT 480
QY 481 GACAAAAAGACACTATTAATTAAGACAGCTGAATGGGGGCTATATCTGCCCTTGA 540
Db 481 GACAAAAAGACACTATTAATTAAGACAGCTGAATGGGGGCTATATCTGCCCTTGA 540
QY 541 AAACGTCTGTCCCAAAATGGGAAAAATTCATAGTTACAATTTCATCTACATCTCT 600
Db 541 AAACGTCTGTCCCAAAATGGGAAAAATTCATAGTTACAATTTCATCTACATCTCT 600
QY 601 TTGGAATTTACCATCAGAGAGCTGAGTGGTGAAGAGATCTGAGAAATGGTGGAGAA 660
Db 601 TTGGAATTTACCATCAGAGAGCTGAGTGGTGAAGAGATCTGAGAAATGGTGGAGAA 660
QY 661 AAGTGGGATCTCCAGATGATCAGAAACATTTGATAGAAAGATTCATTTATAGAGAA 720
Db 661 AAGTGGGATCTCCAGATGATCAGAAACATTTGATAGAAAGATTCATTTATAGAGAA 720
QY 721 TGGTCAAAAAAGACAAAGACCAATTTACATAGTGTGAGTTGGTGCCTACAGAAAA 780
Db 721 TGGTCAAAAAAGACAAAGACCAATTTACATAGTGTGAGTTGGTGCCTACAGAAAA 780
QY 781 GACCTGATCAAGATTAATATGAGTGTGAGTTGGTGCCTACAGAAAA 840
Db 781 GACCTGATCAAGATTAATATGAGTGTGAGTTGGTGCCTACAGAAAA 840
QY 841 TGGAGCTGGGCACTACGGAATTTGTTCGGTTTGGTGTATGATCTGAGAAAA 900
Db 841 TGGAGCTGGGCACTACGGAATTTGTTCGGTTTGGTGTATGATCTGAGAAAA 900
QY 901 ACCTGGAATTAAGATCTTTTGAAGCTTTTAATAGAGAGATGATGATTA 954
Db 901 ACCTGGAATTAAGATCTTTTGAAGCTTTTAATAGAGAGATGATGATTA 954

RESULT 4
LOCUS AE001813 10244 bp DNA linear BCT 02-JUN-1999
DEFINITION Thermotoga maritima section 125 of 136 of the complete genome.
ACCESSION AE001813 AE000512
VERSION AE001813.1 GI:4982321
KEYWORDS
SOURCE Thermotoga maritima.
ORGANISM Thermotoga maritima; Thermotogales; Thermotogaceae;
Bacteria; Thermotoga maritima.
REFERENCE
AUTHORS 1 (bases 1 to 10244)
Neelson, K.E., Clayton, R.A., Gill, S.R., Gwinn, M.L., Dodson, R.J.,
Hart, D.H., Hickey, E.K., Peterson, J.D., Nelson, W.C., Ketchum, K.A.,
McDonald, L., Utterback, T.R., Malek, J.A., Linher, K.D., Garrett, M.M.,
Stewart, A.M., Cotton, M.D., Pratt, M.S., Phillips, C.A.,
Richardson, D., Heidelberg, J., Sutton, G.G., Fleischmann, R.D.,
Eisen, J.A., Fraser, C.M. et al.
Evidence for lateral gene transfer between Archaea and bacteria
from genome sequence of Thermotoga maritima
Nature 399 (6734), 323-329 (1999)
JOURNAL
MEDLINE 99287316
PUBMED 10360571
REFERENCE 2 (bases 1 to 10244)
Neelson, K.E., Clayton, R.A., Gill, S.R., Gwinn, M.L., Dodson, R.J.,
Hart, D.H., Hickey, E.K., Peterson, J.D., Nelson, W.C., Ketchum, K.A.,
McDonald, L., Utterback, T.R., Malek, J.A., Linher, K.D., Garrett, M.M.,
Stewart, A.M., Cotton, M.D., Pratt, M.S., Phillips, C.A.,
Richardson, D., Heidelberg, J., Sutton, G.G., Fleischmann, R.D.,
White, O., Salzberg, S.L., Smith, H.O., Venter, J.C. and Fraser, C.M.
Direct Submission
Submitted (01-JUN-1999) The Institute for Genomic Research, 9712
Medical Center Dr, Rockville, MD 20850, USA
TITLE
JOURNAL
FEATURES
Source location/Qualifiers
1..10244
/organism="Thermotoga maritima"
/db_xref="taxon:2336"
73..1104
/gene="TM1744"
73..1104
/note="similar to GB:U00096 SP:P75823 PID:1787095
PID:2225892 percent identity: 67.58; identified by
sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="L-allo-threonine aldolase"
/protein_id="AAD36809.1"
/db_xref="GI:4982322"
/translation="MIDLSQVYTRKTEEMRKMAQAEVGDVYGEPTINELRLAA
ETFGKALFVSGTGNQVSIAMHQRGEVLEADSHIFRYEGAMVLSGVMPHP
VGGKNGAMPDDVRKAIIRPNHIFPTSLIAENTNRSGRVPLENKEITIKKE
HGIVNIDGARIFNASIAGVADSVFCLSGKCAPSVVAGDDETER
ARKARKMLGGMGAGVLAAGIITLTKWDRKEDHENARLALKIKELGYSVNE
VTNNVILIRDTNKKVNAHGFIETLRNSGVLANVSDTEIRLVTGHKYSRNDIEALNI
FEKLFRRKFS
1183..1806
/gene="TM1745"
1183..1806
/note="similar to percent identity: 0.00; identified by
sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAD36810.1"
/db_xref="GI:4982323"
/translation="MKRFVVMVLITVALGVFAAGYGAERPAFERPQMGVQAQFPG
CGGGLGIGPNAPMYERQPMARERGGFGLQIIRDEQLAKTRIAELNNLNDLD
DQIKELTVAKETREQLSLKREKVOJHEEYNNALYKRDTKTAEKLDQIDDLRDEVQ
DIYKDFADKVDIITVDYRRKFGDKRMFLVLTDEGFVLEENVOE"

gene 2074..4002
CDS /gene="TM1746"
2074..4002
/gene="TM1746"
/note="similar to GB:AE000783 percent identity: 56.54;
identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="oligopeptide ABC transporter, periplasmic
oligopeptide-binding protein"
/protein_id="AAD36811.1"
/db_xref="GI:4982324"
/translation="MKRFVGMVLVLAIFYSEVEVQLPMEGVTFNDPYVQAQHL
ACQTFEGELGVDPKSLGKVVPAVCESTVEDGILTYFTYLRKDRKRSNGDPV
AYPPEFAMKRAASPEPTLPIMASPVQYIKYNDCKSGAKPLDELGVKALDYLQVY
LSRPMSFTIMLVLGAMPLEHRTVOEHEDMMKPEYFNGPYIESFTNVEIYV
RNNYVDEPFGNDRKLVKGGGIGLQGYIAGEIDAVETAVGVDFALKNQSLSELH
EESGVWVGVEITRSLSPYEDDIRIKALAMAIKRVLTIDYVLSGMALPTNACSPDS
EIEAVVGIPIYDEQAKKLAEAGYPNGKGFPRVRYITGASDPVAELVDQMKVVG
ITFEIENESGQSTLQWSDPYVWAPGVCITIGPVMQDSGLMQTDHFTWYDYP
AEVRKLYELREKRESYLREGKTEEMKEILNRAQELMEKQVNSVEDPEFLKRY
FLPAPBELTFPIQIKLYEEMKATTDQKIQLMRLAMRALIYNOEYSIVETGNMPT
NREARKLMARILSPYDKAVEYAKKIQQLQVLDLAYEVLVQKVVYLAAPYKGIYVY
KFSWGPTEFLFKTLTYEK"
3999..5024
/gene="TM1747"
3999..5024
/gene="TM1747"
/note="similar to GB:LA2023 SP:P45054 PID:1006439
PID:1221243 PID:1205369 percent identity: 65.90;
identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="oligopeptide ABC transporter, permease protein"
/protein_id="AAD36812.1"
/db_xref="GI:4982325"
/translation="MNSPDPFGILLNMSGVDMRFLKRLKLLTLTIAISVVYV
ITYVLMVLAPENFEFLQVRDALARVETPPDPAYQATKKEFERYGLNIPKQILMY
LKGAIVFERGFSFSDPARNIEDLKEFTFTFLAUSLILFALVGVPIGLIALAKN
TWIDYATNVSVIGVAIPSYVVAIFLLISYIVGLWIPSGWGIIRTKLPTALALG
PLASVARTRVSLIDTLNODFIRVAVKGGDDRTVIMKIALRSPMPLVYIVGPQAY
LMGVVWVENIFRIPLGQLFANAARVDYPLVTSFTLALTVMIMNLIVDLVALIL
DPRIKD"
5042..5911
/gene="TM1748"
5042..5911
/gene="TM1748"
/note="similar to GB:M57689 SP:P24139 GB:X56347 PID:143606
PID:40007 percent identity: 64.69; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/product="oligopeptide ABC transporter, permease protein"
/protein_id="AAD36813.1"
/db_xref="GI:4982326"
/translation="MLRKNKTESIGWKAPFLRFKKNKMAVIGGVFLIILALAILAP
VTAPVPDEPHIRAFDESPKDFIFGDALGRDLPSTLISLRACILIGSGOVVLI
IGGILGAAGFERGWDIKFIMSIDVDMFAETFTFENVILYATLARGIFLIALGLG
WAGMARLIVGQVLYLKNSEFVGNKAACASFTYIRKHILNMLGPTLVNLASVPRA
NMTESGLAIVIGMVGVRPMPMSWGNLIGEGIMMAFPHLLIFPAVTFATLTISFVLAD
GLRDAFNRSEIS"
5908..6882
/gene="TM1749"
5908..6882
/gene="TM1749"
/note="similar to SP:P04285 GB:X05491 GB:X52093 PID:47805
percent identity: 69.65; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/product="oligopeptide ABC transporter, ATP-binding
protein"
/protein_id="AAD36814.1"


```
/db_xref="GI:4982327"
/translation="MMELNINNLKVEFRHVEGIKAVDGIYSKLKNGSGIIVGESG
SGKSVSLILRLINRGRIYDGEALIFLCKDLKLNKEPRLIRKDSIIITONMITS
LNPIIRVQIOWNEPIIWHRLMKNEARENAITLLEVEGIPSEPKPLNLPFGSGMR
QRMVIMALACHPKLLIADDEPTALDVTIOAOIMELDLOLKEYSNYSVFIETHJSA
NFCORIIITMYAGKIVEAPVEIKTLPHPYTKLSTLEIGSGKGLVPIPNPND
NPTKHPSCGKFEHRCFSFAMEICQREEPVLNISENHRVACHLKGESK"
6852..7838
/gene="TM1750"
/length=986
/locus_tag="TM1750"
/notes="similar to SP:P42065 PID:677944 GB:AL009126 percent
identity: 76.43; identified by sequence similarity;
putative"
/codon_start=1
/transl_table=11
/product="oligopeptide ABC transporter, ATP-binding
protein"
/protein_id="AAD36815.1"
/db_xref="GI:4982328"
/translation="MPDGGGIMKRPILQTVDLKKYFPGKRILKAVDGISTEIKEGE
TLGIYGESGCKSTIGRTILKLRPDGKIFREGDITNLNKEKPKRKQIIFOD
PLGLNPQIVTGRITIEDPLIHKIGTKERKRVBELDMVGIGREFINSFHEFSG
QOORIGIARALALNPKFIVCEPVALDVSIQAOIIDLEEQKMGISYLFIAHNL
VEHISHKVAVMYLGKIVEGVDVKIFLPHIPYRALIKSVKIPMDQOKRFSYSLK
GELPSPIDLPKRCFRQTRCTEKKAIICEKEPELVEKHNHVSCHLVRSRG"
7898..8851
/gene="TM1751"
/length=953
/locus_tag="TM1751"
/notes="similar to SP:P23340 percent identity: 52.35;
identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="endoglucanase"
/protein_id="AAD36816.1"
/db_xref="GI:4982329"
/translation="MGVDPFERNKILRGINIGALNLEAPNEGDMGVILKEPFDIIE
AGEFHRIPRIMSTHAVAPEPYKIMDRFQRVDEVINGLKGALVAVINIHYEELM
DPEHKERFIALMKQIADRYKDPETLEFELIENHNGITLTPKKNELLLEAKVRSI
DKKHTIITIGTAWGSIISALEKISVPEKESNIVTTHYNPEFTIQCAGWEGSKWL
GKRWGSPDDKHLIEEFNFIEMSKRNKRPYIGFGAARKDLERSRIKWSFVVEREM
EKRRMSMAVWECGSGFVYDILRTKWNLDLALIGDSIE"
8844..9833
/gene="TM1752"
/length=989
/locus_tag="TM1752"
/notes="similar to SP:P23340 percent identity: 51.51;
identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="endoglucanase"
/protein_id="AAD36817.1"
/db_xref="GI:4982330"
/translation="NNNTIIPWRGFNLLEAFSIKSTGNFKEEDFLMMAQWDFNVRIP
MCHLMSDRGNPFIIREDFEKIDRIVFGEKYGJHICISLHARAGSYNKEVEKTN
LWKDETQAEAFIHMSFIARRYKGISISHLNINPEPEPQIMASVEDHNSLIRKT
ITEIRKIDPERLIIDIGYGNIPYVDLITENTVQSCGIYIFSTVTHKAENVDSKDF
```

Query Match 94.7%; Score 903; DB 1; Length 10244;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 953; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
QY 1 ATGGGTGTTGATCCTTTGAAGAACAAATATTGGGAAGGCTTAATATAGCAAT 60
DB 7898 ATGGGTGTTGATCCTTTGAAGAACAAATATTGGGAAGGCTTAATATAGCAAT 7957
QY 61 GCGCTTGAAGCAACAATGAGGAGACTGGGAGTGTGATATAAAGATAGTCTTCGAC 120
DB 7958 GCGCTTGAAGCAACAATGAGGAGACTGGGAGTGTGATATAAAGATAGTCTTCGAC 8017
QY 121 ATTATTAAGAGCCGTTTCTCATGTTGCAATTCAATAAGATGAGTACGACGCT 180
```

```
DB 8018 ATTATTAAGAGCCGTTTCTCATGTTGCAATTCAATAAGATGAGTACGACGCT 8077
QY 181 TAGCGTTTCTCTTATTAATCATGATCGCTTCTTCAAAAGATGATGAAGTATA 240
DB 8078 TAGCGTTTCTCTTATTAATCATGATCGCTTCTTCAAAAGATGATGAAGTATA 8137
QY 241 AACGAGCCCTGAAGAGAGACTGGCTGTGTATATAATATCATCTCTGAGGAGTTA 300
DB 8138 AACGAGCCCTGAAGAGAGACTGGCTGTGTATATAATATCATCTCTGAGGAGTTA 8197
QY 301 ATGAATGATCCAGAAAGACACAAAGAAAGATTTCTGCTCTTTGGAACAAATTCGTAT 360
DB 8198 ATGAATGATCCAGAAAGACACAAAGAAAGATTTCTGCTCTTTGGAACAAATTCGTAT 8257
QY 361 CGTTATTAAGACTATCCCGAACTCTATTTTGAATTTGTAATGACCTCAGCGAAT 420
DB 8258 CGTTATTAAGACTATCCCGAACTCTATTTTGAATTTGTAATGACCTCAGCGAAT 8317
QY 421 CTACTCCGGAAGAAATGGAATGAACTGCTTGAAGAGCTTAAAGTATTAAGATCAATT 480
DB 8318 CTACTCCGGAAGAAATGGAATGAACTGCTTGAAGAGCTTAAAGTATTAAGATCAATT 8377
QY 481 GACAAAAGCACACTATATATTAGGCACAGCTGAATGGGGGTATATCTGCCCTTGA 540
DB 8378 GACAAAAGCACACTATATATTAGGCACAGCTGAATGGGGGTATATCTGCCCTTGA 8437
QY 541 AACTGCTGCTCCCAAAATATGGGAAAAATTTCTATAGTTTCAATTCTACTCAATCCT 600
DB 8438 AACTGCTGCTCCCAAAATATGGGAAAAATTTCTATAGTTTCAATTCTACTCAATCCT 8497
QY 601 TTGCAATTTACCATCAGAGAGCTGAGTGGGTGAAGATCTGAGAAATGTTGGGAGA 660
DB 8498 TTGCAATTTACCATCAGAGAGCTGAGTGGGTGAAGATCTGAGAAATGTTGGGAGA 8557
QY 661 AACTGGGATCTCCAGATGATGCAAGAAACATTTGATAGAAATTCATTTTATAGAAGA 720
DB 8558 AACTGGGATCTCCAGATGATGCAAGAAACATTTGATAGAAATTCATTTTATAGAAGA 8617
QY 721 TGCTCAAAAAGAAACAAACCAATTTACATAGTGAGTGGTGGCTACAGAAAAGCT 780
DB 8618 TGCTCAAAAAGAAACAAACCAATTTACATAGTGAGTGGTGGCTACAGAAAAGCT 8677
QY 781 GACCTTGAATCAAGATTAATAATGACCTCTTGTGCTGGCAGAAATGAGAAAAGAGA 840
DB 8678 GACCTTGAATCAAGATTAATAATGACCTCTTGTGCTGGCAGAAATGAGAAAAGAGA 8737
QY 841 TGGAGCTGGGCATGACTGGAATTTGTCGGTGGTGGTGGTATGATCTGAGAAAA 900
DB 8738 TGGAGCTGGGCATGACTGGAATTTGTCGGTGGTGGTGGTATGATCTGAGAAAA 8797
QY 901 ACCTGGAATTAAGATCTTTAGAGCTTTATAGAGGAGATAGCAATTAATAA 954
DB 8798 ACCTGGAATTAAGATCTTTAGAGCTTTATAGAGGAGATAGCAATTAATAA 8851
```

RESULT 5 AC079067 150882 bp DNA linear HTG 18-NOV-2000

DEFINITION Homo sapiens chromosome 4 clone RP11-77P11 map 4, WORKING DRAFT

ACCESSION AC079067

VERSION AC079067.2 GI:11225393

KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.

SOURCE Homo sapiens.

ORGANISM Homo sapiens.

REFERENCE 1 (bases 1 to 150882) Mammalia: Eutheria: Primates: Catarrhini: Hominoidea: Homo.

AUTHORS Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N.,

TITLE
JOURNAL
COMMENT

Anderson, S., Barna, N., Bastien, V., Beda, F., Boguslavskiy, L.,
Boukhgalter, B., Brown, A., Burkett, G., Campopiano, A., Castile, A.,
Choepe, X., Colangelo, M., Collins, S., Collymore, A., Cooke, P.,
DeRellano, K., Dewar, K., Diaz, J. S., Dodge, S., Ferreira, P.,
Fitzhugh, W., Gage, D., Galagan, J., Gardyna, S., Ginde, S., Goyette, M.,
Graham, L., Grand-Pierre, N., Hagos, B., Heatford, A., Horton, L.,
Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A., Lahocque, K.,
Lamarez, R., Landers, T., Lehoczy, J., Levine, R., Lieu, C., Liu, G.,
Macdonald, P., Margus, N., McCarthy, M., McEwan, P., McKernan, K.,
McHeeters, R., Meldrum, J., Menus, L., Mhova, T., Mlenga, V.,
Morro, J., Murphy, T., Naylor, J., Norman, C. H., O'Connor, T.,
O'Donnell, P., O'Neill, D., Oliver, T. M., Oliver, J., Peterson, K.,
Pierre, N., Pisan, C., Pollara, V., Raymond, C., Rieback, M., Riley, R.,
Rogov, P., Rothman, D., Roy, A., Santos, R., Schauer, S., Severy, P.,
Sougnaz, C., Spencer, B., Stange-Thomann, N., Stojanovic, N.,
Strauss, N., Subramanian, A., Talamas, J., Testaye, S., Theodore, J.,
Tirrell, A., Travers, M., Trigglio, J., Vassiliev, H., Viel, R., Vo, A.,
Zimmer, B., Wu, X., Wyman, D., Ye, W. J., Young, G., Zainoun, J.,
Zimmer, A. and Zody, M.

Submitted (17-AUG-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Nov 18, 2000 this sequence version replaced gi:9837996.
All repeats were identified using RepeatMasker:
http://ftp.genome.washington.edu/M/RepeatMasker.html

Center: Whitehead Institute/ MIT Center for Genome Research
Genome Center

Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence.submissions@genome.wi.mit.edu

Project Information

Center project name: L10624

Center clone name: 77_P-11

Summary Statistics

Sequencing vector: plasmid; n/a; 100% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.960731

Consensus quality: 137969 bases at least Q40

Consensus quality: 144918 bases at least Q30

Consensus quality: 147617 bases at least Q20

Insert size: 157000; agarose-fp

Insert size: 148982; sum-of-contigs

Quality coverage: 3.8 in Q20 bases; sum-of-contigs

Quality coverage: 4.0 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 20 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 12294: contig of 12294 bp in length
* 12295 12394: gap of 100 bp
* 12395 14197: contig of 1803 bp in length
* 14198 14297: gap of 100 bp
* 14298 43602: contig of 29305 bp in length
* 43603 43702: gap of 100 bp
* 43703 44846: contig of 1144 bp in length
* 44847 44946: gap of 100 bp
* 44947 46353: contig of 1407 bp in length
* 46354 46453: gap of 100 bp
* 46454 48052: contig of 1599 bp in length
* 48053 48152: gap of 100 bp
* 48153 51116: contig of 2964 bp in length
* 51117 51216: gap of 100 bp
* 51217 52873: contig of 1657 bp in length
* 52874 52973: gap of 100 bp
* 52974 56008: contig of 3035 bp in length
* 56009 56108: gap of 100 bp
* 56109 59527: contig of 3419 bp in length

FEATURES
source

* 59528 59627: gap of 100 bp
* 59628 62795: contig of 3168 bp in length
* 62796 62895: gap of 100 bp
* 62896 67332: contig of 4437 bp in length
* 67333 67432: gap of 100 bp
* 67433 72873: contig of 5441 bp in length
* 72874 72973: gap of 100 bp
* 72974 81642: contig of 8669 bp in length
* 81643 81742: gap of 100 bp
* 81743 87826: contig of 6084 bp in length
* 87827 87926: gap of 100 bp
* 87927 98128: contig of 10202 bp in length
* 98129 98228: gap of 100 bp
* 98229 109128: contig of 10900 bp in length
* 109129 109228: gap of 100 bp
* 109228 125996: contig of 16768 bp in length
* 125997 126096: gap of 100 bp
* 126097 149933: contig of 23837 bp in length
* 149934 150033: gap of 100 bp
* 150034 150882: contig of 849 bp in length.

Location/Qualifiers

1. 150882

/organism="Homo sapiens"

/db_xref="taxon:9606"

/chromosome="4"

/map="4"

/clone="Rp11-77p11"

/clone_lib="RPC1-11 Human Male BAC"

1. 12294

/note="assembly-fragment"

clone_end:56

vector_side:left"

12395. 14197

/note="assembly-fragment"

14298. 43602

/note="assembly-fragment"

43703. 44846

/note="assembly-fragment"

44947. 46353

/note="assembly-fragment"

46454. 48052

/note="assembly-fragment"

48153. 51116

/note="assembly-fragment"

51217. 52873

/note="assembly-fragment"

52974. 56008

/note="assembly-fragment"

56109. 59527

/note="assembly-fragment"

59628. 62795

/note="assembly-fragment"

62896. 67332

/note="assembly-fragment"

67433. 72873

/note="assembly-fragment"

72974. 81642

/note="assembly-fragment"

81743. 87826

/note="assembly-fragment"

87927. 98128

/note="assembly-fragment"

98229. 109128

/note="assembly-fragment"

109229. 125996

/note="assembly-fragment"

126097. 149933

/note="assembly-fragment"

150034. 150882

/note="assembly-fragment"

/note="assembly-fragment"

vector_end:right"

BASE COUNT 46319 a 27306 c 26924 g 48428 t 1905 others

ORIGIN

Query Match 2.4%; Score 23; DB 2: Length 150882;
 Best Local Similarity 100.0%; Pred. No. 0.4;
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 563 AAAAAATCTATAGTTACATT 585
 ||||||||||||||||||
 Db 5557 AAAAAATCTATAGTTACATT 5579

RESULT 6

AC092659 160529 bp DNA linear PRI 01-MAR-2002
 LOCUS Homo sapiens BAC clone RP11-495115 from 4, complete sequence.
 AC092659 AC027044
 AC092659.2 GI:16303561
 HTG.
 KEYWORDS Homo sapiens.
 SOURCE Homo sapiens.
 ORGANISM Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.

REFERENCE
 AUTHORS Sulston, J.E. and Waterston, R.
 TITLE Toward a complete human genome sequence
 JOURNAL Genome Res. 8 (11), 1097-1108 (1998)
 MEDLINE 99063792
 PUBMED 9847074

REFERENCE
 AUTHORS Martin, S., Haakenson, W., and Elliott, G.
 TITLE The sequence of Homo sapiens BAC clone RP11-495115
 JOURNAL Unpublished (2001)
 REFERENCE
 AUTHORS Waterston, R.H.
 TITLE Direct Submission
 JOURNAL Submitted (19-JUL-2001) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA

REFERENCE
 AUTHORS Waterston, R.H.
 TITLE Direct Submission
 JOURNAL Submitted (20-OCT-2001) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA

REFERENCE
 AUTHORS Waterston, R.H.
 TITLE Direct Submission
 JOURNAL Submitted (23-OCT-2001) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA

REFERENCE
 AUTHORS Waterston, R.
 TITLE Direct Submission
 JOURNAL Submitted (01-MAR-2002) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
 On Oct 20, 2001 this sequence version replaced gi:14916243.

COMMENT

Center: Washington University Genome Sequencing Center
 Center code: WUGSC
 Web site: <http://genome.wustl.edu/gsc>
 Contact: saplens@wustl.edu
 Summary Statistics
 Center Project name: H_NH0495115
 Drafting Center: WIBR

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
 all regions were double stranded, sequenced with an alternate

chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:
 Mapping information for this clone was provided by Dr. John D. McPherson, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION:
 The RP11-11 human BAC library was made from the blood of one male donor, as described by Osoegawa, K., Woon, P.Y., Zhao, B., Frengen, E., Tatenno, M., Catanese, J.J., and de Jong, P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pletter de Jong and coworkers at <http://www.chori.org>
 VECTOR: pBACe3.6

NEIGHBORING SEQUENCE INFORMATION:
 The clone sequenced to the left is RP11-721f3; the clone sequenced to the right is AC079067. Actual start of this clone is at base position 1 of RP11-495115; actual end is at base position 160529 of RP11-495115.

The regions between bases 1 to 110 and 659 to 706 are represented only by PCR products of clone DNA. The region between 160353 to 160424 is covered only by a single plasmid subclone.

The sequence of AC027044 has been incorporated into AC092659.

FEATURES

source	Location/Qualifiers
1.160529	/organism="Homo sapiens"
	/db_xref="taxon:9606"
	/chromosome="4"
	/map="4"
	/clone="RP11-495115"
	/clone_1lb="RP11-11"
repeat_region	575..636
	/rpt_family="ERYL"
repeat_region	657..707
	/rpt_family="(GA)n"
repeat_region	701..1179
	/rpt_family="ERYL"
repeat_region	2817..2856
	/rpt_family="L2"
repeat_region	3473..3786
	/rpt_family="MER1_type"
repeat_region	3595..3650
	/rpt_family="AT_rich"
repeat_region	4674..5020
	/rpt_family="Alu"
repeat_region	4769..4814
	/rpt_family="A-rich"
repeat_region	6058..6353
	/rpt_family="ERYL"
repeat_region	6731..6773
	/rpt_family="AT_rich"
repeat_region	6834..6859
	/rpt_family="AT_rich"
repeat_region	7045..7078
	/rpt_family="AT_rich"
repeat_region	7795..7848
	/rpt_family="AT_rich"
repeat_region	7942..7972
	/rpt_family="AT_rich"
repeat_region	8306..8333
	/rpt_family="AT_rich"
repeat_region	8605..8638
	/rpt_family="(TTTTC)n"

repeat_region 8917. .8941
/rpt_family="(T)n"
repeat_region 10113. .10421
/rpt_family="Alu"
repeat_region 10398. .10421
/rpt_family="(A)n"
repeat_region 10448. .10481
/rpt_family="AT-rich"
repeat_region 11924. .11959
/rpt_family="(TAAA)n"
repeat_region 12981. .13071
/rpt_family="MIR"
repeat_region 13256. .13326
/rpt_family="(TA)n"
repeat_region 13451. .13484
/rpt_family="MIR"
repeat_region 14420. .14543
/rpt_family="L1"
repeat_region 14615. .14725
/rpt_family="L1"
repeat_region 14808. .14913
/rpt_family="MIR"
repeat_region 15012. .15040
/rpt_family="AT-rich"
repeat_region 15031. .15067
/rpt_family="L1"
repeat_region 15110. .15201
/rpt_family="L2"
repeat_region 15849. .15927
/rpt_family="L2"
repeat_region 18299. .18599
/rpt_family="Alu"
repeat_region 18821. .18859
/rpt_family="AT-rich"
repeat_region 18979. .19430
/rpt_family="L2"
repeat_region 19489. .19620
/rpt_family="(CATA)n"
repeat_region 19659. .19717
/rpt_family="(CATATA)n"
repeat_region 21247. .21291
/rpt_family="L2"
repeat_region 21468. .21624
/rpt_family="L1"
repeat_region 21665. .22429
/rpt_family="L1"
repeat_region 22430. .22740
/rpt_family="Alu"
repeat_region 22712. .22757
/rpt_family="A-rich"
repeat_region 22741. .24033
/rpt_family="L1"
repeat_region 24034. .24319
/rpt_family="L1"
repeat_region 24320. .25563
/rpt_family="L1"
repeat_region 24988. .25014
/rpt_family="(TAAAA)n"
repeat_region 25576. .25988
/rpt_family="L1"
repeat_region 26091. .27707
/rpt_family="L1"
repeat_region 27719. .27838
/rpt_family="L1"
repeat_region 27839. .27873
/rpt_family="MALR"
repeat_region 27874. .28172

Query Match 2.4%; Score 23; DB 9; Length 160529;
Best Local Similarity 100.0%; Pred. No. 0.4;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 563 AAAAAATCTATAGTTACAAAT 585

Db 122669 AAAAAATCTATAGTTACAAAT 122691

RESULT 7

AC126611/c 178997 bp DNA linear HTG 07-JUL-2002
LOCUS Mus musculus chromosome UNK clone RP23-316N22, WORKING DRAFT
DEFINITION SEQUENCE, 10 unordered pieces.

ACCESSION

AC126611 GI:21702974
AC126611.1
HTG: HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
KEYWORDS house mouse.

SOURCE

Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE

1 (bases 1 to 178997)
McPherson,J.D. and Waterston,R.H.
The sequence of Mus musculus clone

JOURNAL

2 (bases 1 to 178997)
McPherson,J.D. and Waterston,R.H.
Direct Submission
Submitted (07-JUL-2002) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA

COMMENT

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: <http://genome.wustl.edu/gsc/index.shtml>
Contact: submissions@wustl.wustl.edu
Project Information
Center project name: M_BA0316N22

Summary Statistics

Sequencing vector: M13: 0%
Sequencing vector: Plasmid: 100%
Chemistry: Dye-terminator Big Dye: 100% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 175866 bases at least Q40
Consensus quality: 176355 bases at least Q30
Consensus quality: 176586 bases at least Q20

* NOTE: This is a 'working draft' sequence. It currently
* consists of 10 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 1844: contig of 1844 bp in length
* 1845 1944: gap of unknown length
* 1945 3071: contig of 1127 bp in length
* 3072 3171: gap of unknown length
* 3172 5257: contig of 2086 bp in length
* 5258 5357: gap of unknown length
* 5358 9354: contig of 3997 bp in length
* 9355 9454: gap of unknown length
* 9455 13669: contig of 4215 bp in length
* 13670 13770: gap of unknown length
* 13770 26976: contig of 13207 bp in length
* 26977 27076: gap of unknown length
* 27077 52620: contig of 25544 bp in length
* 52621 52720: gap of unknown length
* 52721 78960: contig of 26240 bp in length
* 78961 79060: gap of unknown length
* 79061 114898: contig of 35838 bp in length
* 114899 114998: gap of unknown length
* 114999 178997: contig of 63999 bp in length.
Location/Qualifiers
1. .178997

FEATURES

source

```

/misc_feature /organism="Mus musculus"
/db_xref="taxon:10090"
/chromosome="X"
/clone="RP23-316N22"
1.1844
/misc_feature /note="assembly_name:Contig11"
1945.3071
/misc_feature /note="assembly_name:Contig12"
3172.5257
/misc_feature /note="assembly_name:Contig13"
5358.9354
/misc_feature /note="assembly_name:Contig14"
9455.13669
/misc_feature /note="assembly_name:Contig15"
13770.26976
/misc_feature /note="assembly_name:Contig16"
27077.52620
/misc_feature /note="assembly_name:Contig17"
52721.78960
/misc_feature /note="assembly_name:Contig18"
79061.114898
/misc_feature /note="assembly_name:Contig19"
114999.178997
BASE COUNT 53895 a 34886 c 33994 g 55211 t 1011 others
ORIGIN

```

```

Query Match 2.4% Score 23; DB 2; Length 178997;
Best Local Similarity 100.0%; Pred. No. 0.4;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 719 AATGTCACAAAGACACAAAGA 741
Db 90950 AATGTCACAAAGACACAAAGA 90928

```

```

RESULT 8 AC108661 194378 bp DNA linear HTG 13-JUL-2002
LOCUS AC108661
DEFINITION Rattus norvegicus clone CH230-231H7, *** SEQUENCING IN PROGRESS
ACCESSION AC108661
VERSION AC108661.3 GI:21737936
KEYWORDS HTG; HTGS; PHASE1.
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.

```

```

REFERENCE 1 (bases 1 to 194378)
AUTHORS Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
Alpbrooks,S.L., Amaralunge,H.C., Are,J.R., Ayale,M., Banks,T.,
Barbieri,J., Benton,J., Bimge,K., Blankenburg,K., Bonnin,D.,
Bouch,J., Bowle,S., Brileva,M., Brown,E., Brown,M., Bryant,N.P.,
Burton,C., Burch,P., Burkett,C., Butrell,K.L., Byrd,N.C.,
Carroll,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
Earhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
Gabriel,J., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
Gorell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
Harris,C., Harris,K., Hart,M., Haylak,P., Hawes,A., Hernandez,J.,
Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B.,
Horn,J., Howard,S., Huber,J., Hulik,S., Hume,J., Jackson,L.E.,
Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
Karlssoon,B., Kelly,S., Khan,U., King,L., Korvach,J., Kovar,C.,
Kratovic,J., Kureishi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
Li,J., Li,Z., Lichtarge,O., Liu,C., Liu,J., Liu,W., Loussegod,H.,
Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,

```

```

TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

```

```

Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
Massey,E., Mawhney,E., McLeod,M.P., Meador,M., Mel,G., Metzger,M.,
Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,M., Nguyen,N.,
Nguyen,N., Nickerson,E., Nwokenwo,S., Oguh,M., Okwunonu,G.,
Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,
Rivers,M., Rojas,A., Rojibokan,I., Rolfe,M., Ruiz,S., Savery,G.,
Scherer,S., Scott,G., Shen,H., Shooshari,N., Sisson,I.,
Sodergren,E., Sonake,T., Sparks,A., Stanley,H., Stone,H.,
Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,
Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S.,
Usmani,K., Vasquez,L., Vera,V., Villalón,D., Vinson,R., Wang,Q.,
Wang,S., Ward-Moore,S., Warren,R., Washington,C., Wallington,S.,
Williams,G., Williamson,A., Wlezyk,R., Woodson,S., Worley,K.,
Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
Weinstock,G., and Gibbs,R.
Direct Submission
Unpublished
2 (bases 1 to 194378)
Worley,K.C.
Submitted (31-JAN-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 194378)
Worley,K.C.
Direct Submission
Submitted (13-JUL-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Jul 12, 2002 this sequence version replaced gi:18846857.
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GPZL
Center clone name: CH230-231H7
----- Summary Statistics
Sequencing vector: Plasmid;
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 124895 bases at least Q40
Consensus quality: 129642 bases at least Q30
Consensus quality: 133710 bases at least Q20
-----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 79 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1 1598: contig of 1598 bp in length
* 1599 1698: gap of unknown length
* 1699 3203: contig of 1505 bp in length
* 3204 3303: gap of unknown length
* 3304 4501: contig of 1198 bp in length
* 4502 4601: gap of unknown length
* 4602 5732: contig of 1131 bp in length
* 5733 5832: gap of unknown length
* 5833 7289: contig of 1457 bp in length
* 7290 7389: gap of unknown length
* 7390 8409: contig of 1020 bp in length
* 8410 8509: gap of unknown length
* 8510 9591: contig of 1082 bp in length
* 9592 9691: gap of unknown length
* 9692 11185: contig of 1494 bp in length

```

11186 11285: gap of unknown length
11286 12518: contig of 1233 bp in length
12519 12618: gap of unknown length
12619 13822: contig of 1204 bp in length
13823 13922: gap of unknown length
13923 15081: contig of 1159 bp in length
15082 15181: gap of unknown length
15182 16462: contig of 1281 bp in length
16463 16562: gap of unknown length
16563 17679: contig of 1117 bp in length
17680 17779: gap of unknown length
17780 19196: contig of 1417 bp in length
19197 19296: gap of unknown length
19297 20635: contig of 1339 bp in length
20636 20735: gap of unknown length
20736 21962: contig of 1227 bp in length
21963 22062: gap of unknown length
22063 23138: contig of 1076 bp in length
23139 23238: gap of unknown length
23239 24790: contig of 1552 bp in length
24791 24890: gap of unknown length
24891 26080: contig of 1190 bp in length
26081 26180: gap of unknown length
26181 27612: contig of 1432 bp in length
27613 27712: gap of unknown length
27713 28829: contig of 1117 bp in length
28830 28929: gap of unknown length
28930 30142: contig of 1213 bp in length
30143 30242: gap of unknown length
30243 31838: contig of 1596 bp in length
31839 31938: gap of unknown length
31939 33398: contig of 1460 bp in length
33399 33498: gap of unknown length
33499 34808: contig of 1310 bp in length
34809 34908: gap of unknown length
34909 36063: contig of 1155 bp in length
36064 36163: gap of unknown length
36164 37521: contig of 1358 bp in length
37522 37621: gap of unknown length
37623 39071: contig of 1450 bp in length
39072 39171: gap of unknown length
39172 40179: contig of 1008 bp in length
40180 40279: gap of unknown length
40280 42264: contig of 1985 bp in length
42265 42364: gap of unknown length
42365 43904: contig of 1540 bp in length
43905 44004: gap of unknown length
44005 46127: contig of 2123 bp in length
46128 46227: gap of unknown length
46228 48010: contig of 1783 bp in length
48011 48110: gap of unknown length
48111 49843: contig of 1733 bp in length
49844 49943: gap of unknown length
49944 52463: contig of 2420 bp in length
52464 52463: gap of unknown length
52464 54700: contig of 2237 bp in length
54701 54800: gap of unknown length
54801 56327: contig of 1527 bp in length
56328 56427: gap of unknown length
56428 57929: contig of 1502 bp in length
57930 58029: gap of unknown length
58030 59377: contig of 1348 bp in length
59378 59477: gap of unknown length
59479 61930: contig of 2443 bp in length
61931 62020: gap of unknown length
62021 63992: contig of 1972 bp in length
63993 64092: gap of unknown length
64093 66233: contig of 2141 bp in length
66234 66333: gap of unknown length
66334 68509: contig of 2176 bp in length
68510 68609: gap of unknown length
68610 70025: contig of 1416 bp in length
70026 70125: gap of unknown length

Query Match 2.3% Score 22: DB 2: Length 194378;
Best Local Similarity 100.0%; Pred. No. 1.4;
Matches 22: Conservative 0; Mismatches 0; Indels 0; Gaps 0;
*
Qy 900 AACCTGATAAAGATCTTTA 921
Db 82561 AACCTGATAAAGATCTTTA 82582
|||||
RESULT 9
LOCUS AR077713 51 bp DNA PAT 31-AUG-2000
DEFINITION Sequence 3 from patent US 5962258.
ACCESSION AR077713
VERSION AR077713.1 GI:10004459
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 51)
AUTHORS Mathur,E.J. and Lam,D.E.
TITLE Carboxymethyl cellulase from *Thermotoga maritima*
JOURNAL Patent: US 5962258-A 3 05-OCT-1999;
FEATURES
source Location/Qualifiers
BASE COUNT 14 a 6 c 14 g 17 t
ORIGIN
Query Match 2.2% Score 21: DB 6: Length 51;
Best Local Similarity 100.0%; Pred. No. 7.6;
Matches 21: Conservative 0; Mismatches 0; Indels 0; Gaps 0;
*
Qy 1 ATGGGTTGATCCTTTGAA 21
Db 31 ATGGGTTGATCCTTTGAA 51
|||||
RESULT 10
LOCUS AR096553 51 bp DNA PAT 08-SEP-2000
DEFINITION Sequence 3 from patent US 6008032.
ACCESSION AR096553
VERSION AR096553.1 GI:10025459
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 51)
AUTHORS Mathur,E.J. and Lam,D.E.
TITLE Carboxymethyl cellulase from *Thermotoga maritima*
JOURNAL Patent: US 6008032-A 3 28-DEC-1999;
FEATURES Location/Qualifiers

	source	1..51	/organism="unknown"				
	BASE COUNT	14 a	6 c	14 g	17 t		
	ORIGIN						
	Query Match	2.2%; Score 21; DB 6;	Length 51;				
	Best Local Similarity	100.0%;	Pred. No. 7.6;				
	Matches	21; Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;		
OY	1 ATGGGTGTTGATCCTTTCGAA 21						
Dn	31 ATGGGTGTTGATCCTTTCGAA 51						
RESULT 11							
A157662	ARI57662	51 bp	DNA	linear	PAT 17-OCT-2001		
LOCUS	Sequence 3 from patent US 6245547.						
DEFINITION	ARI57662						
ACCESSION	ARI57662						
VERSION	ARI57662.1 GI:16218625						
KEYWORDS	.						
SOURCE	Unknown.						
ORGANISM	Unknown.						
REFERENCE	Unclassified.						
AUTHORS	1 (bases 1 to 51)						
TITLE	Mathur,E.J. and Lam,D.E.						
JOURNAL	Carboxymethyl cellulose from thermotoga maritima						
FEATURES	Patent: US 6245547-A 3 12-JUN-2001; Location/Qualifiers 1..51						
base count	14 a	6 c	14 g	17 t			
ORIGIN	/organism="unknown"						
Query Match	2.2%; Score 21; DB 6;	Length 51;					
Best Local Similarity	100.0%;	Pred. No. 7.6;					
Matches	21; Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;			
OY	1 ATGGGTGTTGATCCTTTCGAA 21						
Dn	31 ATGGGTGTTGATCCTTTCGAA 51						
RESULT 12							
HS86F14							
LOCUS	HS86F14	106571 bp	DNA	linear	PRI 23-NOV-1999		
DEFINITION	Human DNA sequence from PAC 86F14 on chromosome 1q23-1q24. Contains coagulation factor V, ESTs and STS.						
ACCESSION	Z99572						
VERSION	Z99572.1 GI:2769646						
KEYWORDS	1q23-1q24; Blood coagulation factor; factor V.						
SOURCE	Homo sapiens. Homo sapiens. Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. 1 (bases 1 to 106571) Direct Submission submitted (13-JAN-1998) Chromosome 1 Project Group (http://www.sanger.ac.uk/HGP/Chr1/) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humangenes@sanger.ac.uk Clone requests: clonerequests@sanger.ac.uk on Jan 13, 1998 this sequence version replaced gi:2578147. IMPORTANT: This sequence is not the entire insert of clone 86F14. It may be shorter because we only sequence overlapping sections once, or longer because we arrange for a small overlap between neighbouring submissions. During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variations annotated may not be found in the sequence submission corresponding to the overlapping clone as we submit sequences with only a small overlap as described above.						
COMMENT							

This sequence was generated from part of bacterial clone contigs of human chromosome 1, constructed by the Sanger Centre chromosome 1 mapping group. Further information can be found at <http://www.sanger.ac.uk/Help/Chr1/>. This sequence has been finished according to sequence map criteria as follows. An attempt is made to resolve all sequencing problems, such as compressions and repeats, but not necessarily within known annotated human repeat sequence elements (e.g. Alu). Where the sequence is ambiguous, there is an annotation using the 'unsure' feature key.

The true right end of clone 206d15 is at 104.

The true right end of clone 86F14 is at 106571.

86F14 is from the library Rpec11 constructed at the Roswell Park Cancer Institute by the group of Pieter de Jong.

For further details see <http://bacpac.med.buffalo.edu/>.

FEATURES	Location/Qualifiers
source	1. .106571
	/organism="Homo sapiens"
	/db_xref="taxon:9606"
	/chromosome="1"
	/map="1q23-1q24"
	/clone="RP1-86F14"
	/clone.lib="RPcl-1"
repeat_region	810. .1090
	/note="AluSc repeat: matches 1. .302 of consensus"
repeat_region	1270. .1360
	/note="M1L1E repeat: matches 1. .84 of consensus"
repeat_region	1410. .1587
	/note="M1R1D repeat: matches 139. .323 of consensus"
repeat_region	1575. .1858
	/note="M1L1E repeat: matches 265. .568 of consensus"
repeat_region	1876. .2745
	/note="L1MB5 repeat: matches 921. .1 of consensus"
repeat_region	2597. .2793
	/note="L1 repeat: matches 5390. .5194 of consensus"
repeat_region	2854. .3352
	/note="L1ME1 repeat: matches 687. .170 of consensus"
repeat_region	2971. .3352
	/note="L1ME2 repeat: matches 570. .170 of consensus"
repeat_region	3327. .3559
	/note="MER20 repeat: matches 1. .218 of consensus"
repeat_region	3580. .3873
	/note="AluSc repeat: matches 1. .299 of consensus"
repeat_region	3877. .4060
	/note="L1MC2 repeat: matches 193. .10 of consensus"
repeat_region	6468. .6575
	/note="M1R2 repeat: matches 22. .145 of consensus"
repeat_region	7024. .7238
	/note="AluY repeat: matches 84. .299 of consensus;
	Incomplete repeat"
repeat_region	7834. .8126
	/note="AluSc repeat: matches 2. .295 of consensus"
repeat_region	10054. .10223
	/note="M1R repeat: matches 89. .262 of consensus"
misc_feature	complement(10749. .11106)
	/note="match: STS G05144"
repeat_region	11725. .12081
	/note="THE1C repeat: matches 371. .1 of consensus"
repeat_region	12196. .12497
	/note="AluSc repeat: matches 303. .1 of consensus"
repeat_region	13253. .13367
	/note="M1R2 repeat: matches 4. .127 of consensus"
repeat_region	13398. .15749
	/note="L1 repeat: matches 2577. .4940 of consensus"
repeat_region	16076. .16554
	/note="L1 repeat: matches 4910. .5390 of consensus"
repeat_region	16409. .17294
	/note="L1P2 repeat: matches 1. .891 of consensus"
repeat_region	17808. .19825
	/note="L1 repeat: matches 3338. .5390 of consensus"
repeat_region	19683. .20571
	/note="L1P6 repeat: matches 1. .890 of consensus"
repeat_region	22732. .22827

```

repeat_region 22748..22851 /note="LIMB4 repeat: matches 797..902 of consensus"
/note="LIMB5 repeat: matches 812..922 of consensus"
prim_transcript <25068..25343 /note="match: H61071 H69565"
repeat_region 25302..25381 /note="MIR repeat: matches 49..131 of consensus"
repeat_region 25527..25703 /note="MIR repeat: matches 1..178 of consensus"
repeat_region 25932..26062 /note="MIR repeat: matches 1..178 of consensus"
repeat_region 26055..26264 /note="MIR repeat: matches 18..138 of consensus"
repeat_region 26265..26378 /note="L1 repeat: matches 2971..2759 of consensus"
prim_transcript 27046..27579 /note="LIMB7 repeat: matches 1017..895 of consensus"
/note="match: multiple ESTs; match: R71060 H69028 R82016
R10102 R82280 R82874 H74282 D85329"
CDS 33638..33782,36315..36470,36919..37022,37955..38026,
39019..39135,41033..41212,42726..42936,43904..44140,
49624..49798,53412..56232,57414..57626,59560..59710,
62919..63133,63758..63857,65675..65852,68300..68465,
69764..69985,72271..72414,73672..73884,85339..85461,
95549..95640,99347..99504))
/codon_start=1
/product="factor V"
/protein_id="CAB16748.1"
/db_xref="GI:2769647"
/db_xref="SPTREMBL:O43737"
/translating="MPCGPRMVLVLTGTSVNGCSGTAEAOIRPFVVAAGSIS
XPEPTNSLNVTSFKIVRETEPKKPKPOSTISGLGPTLVAAGDITVH
KKAKRPLSIHQIGIRSKLSGASGLDTPAEKMDVAAGRETTWMSSEBSG
THDDPCLTHIYSHENLIEDNSGLIGLCKKGTLEGGTQKEDQIVLEVF
DEKSSOSSILTVNGVNGTMDPITVCAADHLSMHLGMSGPELFIHNGQVL
RONHKKVSAITLVASTANTVGEKGIISLPRKLGOMAYIDIKNPKTR
NKKITREORRRHMKREYFIAAEVITGAPVAMPADKYSOHLNASSPSYIPGV
KMYTOYEDESEFTSGRNNMTLRAVPGETTYRKNIILEDEPTEMDACLRYSQD
TESPYEDENSSFTSGRNNMTLRAVPGETTYRKNIILEDEPTEMDACLRYSQD
VIMRDIASGLIGLICKSRSLDROGIDRAIDEQAAFAVDEKSNYLEDNIKE
CENPDEKRDDEPKFEYNIMSTINGVPSITLIGCPEDTVQVHFCSTQVNEILTI
HFGHSEFYGKREDITLTFPMRGESVYTMVDVGTWMLTSMNSSPRKLRFEDV
KCTPDDEDSYEIEFEPESTVATRKMDLEDEDEADYQNRKLAALGIRSEF
NSLNDEBEERFNLALANGTEFVSNDIIVGNSYNSNKTFFVNLLEAPQAP
SHQOATTAGSPPLHLGKNSVYNSSTAHSSPSPEDIDPQDVTGTRLSLAGE
FKSQEHAKHKGPRVDOAKHRSMAKLHAGVHLSIDTSPSGMRPMDLPSOD
TSPSRMRPKDPDDLILKQNSSKILLGRWHLSEKSTETIDTDEDTAVNMIL
ISPONASRWAGESTPLANKGQSGHPRVAKSLQVNDGGRKLKSOPLIKTR
KKKEKHTHAPLPRFPHLRSEAYNTESEERLKSILVAKNENSLPDLMOQLPS
MDGCVTASLPHDNQNSNDTGOASCPGLVQVYVPEHQTFFIOPDDHMTSDPSH
RSSPSELSEMLEYDRSHKSPDIDISOMSPSEHVMQVYISPOLSTISPELSONL
SPDLSTPELILORLSPALQMPISPDLSPLSHHTLSIDTSONLSPDL
SQTNLSPALGOMPLSPDLSTLSDLSQTNLSPELSHHTLSIDTSONLSPDL
ISPDLSTLSDLSQTNLSPELSONLSPALQMPISPDLSPLSHHTLSIDTSONLSP
LSQTNLSPDLSEMPILFADLSQILPDLDMQTNLSPDLSEMPILFADLSQILP
TLPDLSDTLTLPDLSDLSQILPDLDMQTNLSPDLSEMPILFADLSQILPDL
SPLTNDSTLTKERNPLVIGLSKGDYIEIIPKEVOSEEDYAEIDYPPDPKPT
DVNTNINSSRDPDNIAMVLRNNGNRRNYIAGEELISVSEVOVRETDIEDSDIP
EDTYYKVRKRLIDSTFTTRDREGREYEHLLGPIIRAEVDVIOVRKNLASRY
SLAHGLSEKSEKGYEDSDSPWEDNAVONNSYTYVMAHTEBSGESGSAAR
AMAYYSANPEKDIHSLGLPLILQKGLIKMMDREVEFLHLTEDEKSKRY
EKKRSRMLTSEEMKSHFHAINGMIVLPGIKMEQWMLNTEVENORAGQDPE
HFGHGTLENGKNOHGVLPPLPGSEKTEMAKSGKMWLLTEVENORAGQDPE
LHNDRCRMGSLGTISDSOIKASEFLIYWEPLRLANNNGSVENNOIMQIFGNSRNV
SKPWIOVDMQEVITITIGTQGAHRYKSCYTFEYVAYASNOIMQIFGNSRNV
YFNGNSDASTIKENQDPPVAVIRISIPRAIRKPIRLLEOCEVNGESTPLGMN
GKIENKOITASSEFKSMWGDYWEFRLAUNQGVNAWQAKANNKKWLEIDLTKIK
ITAITIGCKSLSEMYKSYTIHYSEQYEMKRYRKSXVADKIFEGNNTGKHVKN
FENPITSRPIRVIPTWNSIALRLEFGDII"
repeat_region 27629..27788 /note="AluJo repeat: matches 302..140 of consensus;
incomplete repeat"

```

```

repeat_region 28002..28176 /note="MIR repeat: matches 75..259 of consensus"
repeat_region 29041..29084 /note="MIR2 repeat: matches 103..146 of consensus"
repeat_region 31903..31988 /note="MIR repeat: matches 104..17 of consensus"
repeat_region 34738..35036 /note="MIR repeat: matches 299..1 of consensus"
repeat_region 35051..35542 /note="L1 repeat: matches 3215..3729 of consensus"
repeat_region 36628..36808 /note="MIR repeat: matches 40..235 of consensus"
repeat_region 38333..38451 /note="AluJo repeat: matches 1..134 of consensus;
incomplete repeat"
repeat_region 40215..40272 /note="MIR repeat: matches 1..53 of consensus"
repeat_region 43274..43327 /note="MIR2 repeat: matches 146..93 of consensus"
repeat_region 43456..43548 /note="LIMB8 repeat: matches 1038..944 of consensus"
repeat_region 44951..46774 /note="L1 repeat: matches 3613..5390 of consensus"
repeat_region 46627..46980 /note="L1 repeat: matches 3613..5390 of consensus"
repeat_region 46983..47446 /note="LIMB7 repeat: matches 2..364 of consensus"

Query Match 2.2%; Score 21; DB 9; Length 106571;
Best Local Similarity 100.0%; Pred. No. 5.4;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 786 TGAATCAAGATTAAGTGAC 806
Db 56614 TGAATCAAGATTAAGTGAC 56634
|||||
AC007870 134757 bp DNA linear PRI 12-AUG-1999
LOCUS AC007870
DEFINITION Genomic sequence for Homo sapiens clone 4p6, complete sequence.
ACCESSION AC007870
VERSION AC007870.3 GI:5731403
KEYWORDS HTG.
SOURCE Homo sapiens.
ORGANISM Homo sapiens.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 134757)
Gnoj,L., Haberman,K., Spiegel,L., de la Bastide,M., Dedhia,N.,
Rodriguez,M., Shah,R., Schutz,K., Shekher,M., Spiegel,L., Swaby,I.,
Vill,D. and McCombie,W.R.
Genomic Sequence for Homo sapiens clone 4p6
Unpublished
2 (bases 1 to 134757)
McCombie,W.R.
Direct Submission
Submitted (19-JUN-1999) Lita Annenberg Hazen Genome Sequencing
Center, Cold Spring Harbor Laboratory, 1 Bunglown Road, Cold Spring
Harbor, NY 11724, USA
3 (bases 1 to 134757)
McCombie,W.R.
Direct Submission
Submitted (12-AUG-1999) Lita Annenberg Hazen Genome Sequencing
Center, Cold Spring Harbor Laboratory, 1 Bunglown Road, Cold Spring
Harbor, NY 11724, USA
COMMENT On Aug 12, 1999 this sequence version replaced gi:5306231.
From 107690-107780, the shotgun assembly of 4p6 was single-stranded
with single chemistry. Cloning vector was pBelOBAC11.
FEATURES
source
1..134757
/organism="Homo sapiens"
/db_xref="taxon:9606"

```


/clone="4p6"
/clone.lib="PCR-screenable Research Genetics Library"
BASE COUNT 41775 a 23641 c 25490 g 43851 t
ORIGIN

Query Match 2.2%; Score 21; DB 9; Length 134757;
Best Local Similarity 100.0%; Pred. No. 5.4;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 32 TATTGGAGAGGCATTATA 52
Db 130408 TATTGGAGAGGCATTATA 130388

RESULT 14
AL355177/c 134781 bp DNA linear ROD 29-JUN-2002
LOCUS Mouse DNA sequence from clone RP21-61H21 on chromosome 5, complete
DEFINITION sequence.
ACCESSION AL355177
VERSION AL355177.14 GI:14529794
KEYWORDS HTG.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 134781)

REFERENCE
AUTHORS Philimore,B.
TITLE Direct Submission
JOURNAL Submitted (29-JUN-2002) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humquerry@sanger.ac.uk Clone requests: clonerequests@sanger.ac.uk
On Jun 22, 2001 this sequence version replaced g1:11610998.

COMMENT
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
The following abbreviations are used to associate primary accession
numbers given in the feature table with their source databases:
Em: EMBL; Sw: SWISSPROT; Tr: TrEMBL; Wp: WormPEP; Information
on the WORMPEP database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep RP21-61H21 is
from the RPII-21 Mouse PAC library
constructed by the group of Pieter de Jong.
For further details see http://www.chori.org/bacpac/home.htm

VECTOR: pPAC4
----- Genome Center
Center: UK Medical Research Council
Center code: UK-MRC
Web site: http://mrseq.har.mrc.ac.uk
Contact: mrseq@har.mrc.ac.uk

FEATURES
source
1. 134781
/organism="Mus musculus"
/db_xref="taxon:10090"
/chromosome="5"
/clone="RP21-61H21"
/clone.lib="RPII-21"
BASE COUNT 37151 a 29553 c 30055 g 38022 t
ORIGIN

Query Match 2.2%; Score 21; DB 10; Length 134781;
Best Local Similarity 100.0%; Pred. No. 5.4;

Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 387 ATTTTGAATTCGATGA 407
Db 35844 ATTTTGAATTCGATGA 35824

RESULT 15
AL513082 137154 bp DNA linear PRI 12-JUN-2001
LOCUS Human DNA sequence from clone RPII-229N15 on chromosome Xq24-25,
complete sequence.
DEFINITION complete sequence.
ACCESSION AL513082
VERSION AL513082.10 GI:14422228
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 137154)

REFERENCE
AUTHORS Ashwell,R.
TITLE Direct Submission
JOURNAL Submitted (11-JUN-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk
On Jun 13, 2001 this sequence version replaced g1:14141422.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >=

30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one plasmid subclone or more than one M13 subclone; and the
assembly was confirmed by restriction digest. The following
abbreviations are used to associate primary accession numbers given
in the feature table with their source databases: Em: EMBL; Sw:
SWISSPROT; Tr: TrEMBL; Wp: WormPEP; Information on the WORMPEP
database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
was generated from part of bacterial clone contigs of human
chromosome X, constructed by the Sanger Centre Chromosome X Mapping
Group. Further information can be found at
http://www.sanger.ac.uk/HGP/ChrX
RPII-229N15 is from the library RPII-11.1 constructed by the group
of Pieter de Jong. For further details see
http://www.chori.org/bacpac/home.htm
VECTOR: pBACe3.6

This sequence is the entire insert of clone RPII-229N15 The true
left end of clone RPII-161K18 is at 87431 in this sequence.

FEATURES
source
1. 137154
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="X"
/map="q24-25"
/clone="RPII-229N15"
/clone.lib="RPII-11.1"
repeat_region 1. 1927
/note="LIM1 repeat: matches 4363. 6304 of consensus"
repeat_region 1961. 2667
/note="LIM2 repeat: matches 5353. 6073 of consensus"
repeat_region 2670. 3022
/note="THE1B repeat: matches 1. 364 of consensus"
repeat_region 3023. 3258
/note="LIM2 repeat: matches 6071. 6307 of consensus"
repeat_region 3282. 4586
/note="LIM1 repeat: matches 4643. 5968 of consensus"
misc_feature 4450. 4843

```
/note="match: GSS: Em:AQ116765"
4716..6342
repeat_region
/note="L1M1 repeat: matches 2845..4472 of consensus"
7020..7364
misc_feature
/note="match: STS: Em:G25694"
7179..7565
/note="match: STS: Em:G11556 Em:G26337"
complement(7212..7657)
misc_feature
/note="match: GSS: Em:AQ054060"
7317..7636
misc_feature
/note="match: GSS: Em:A2694362"
7650..8451
repeat_region
/note="L1M1 repeat: matches 4463..5230 of consensus"
8452..8861
repeat_region
/note="MSTB repeat: matches 4..426 of consensus"
8862..9531
repeat_region
/note="L1M1 repeat: matches 5230..5998 of consensus"
9586..9654
repeat_region
/note="L2 repeat: matches 2262..2327 of consensus"
9655..9952
repeat_region
/note="AluSc repeat: matches 3..304 of consensus"
9953..10352
repeat_region
/note="L2 repeat: matches 2327..2750 of consensus"
10508..10587
repeat_region
/note="40 copies 2 mer at 67% conserved"
10826..10972
repeat_region
/note="MER5B repeat: matches 31..178 of consensus"
11890..12256
repeat_region
/note="MER47A repeat: matches 1..365 of consensus"
14740..14799
repeat_region
/note="30 copies 2 mer tt 76% conserved"
14807..16822
repeat_region
/note="L1P43 repeat: matches 4131..6146 of consensus"
16822..19172
repeat_region
/note="L1P43 repeat: matches 1777..4135 of consensus"
21608..21672
repeat_region
/note="MIR repeat: matches 86..148 of consensus"
22071..22375
repeat_region
/note="AluSc repeat: matches 1..305 of consensus"
22850..22877
repeat_region
/note="14 copies 2 mer tg 100% conserved"
23078..23131
repeat_region
/note="27 copies 2 mer ta 74% conserved"
23148..23523
repeat_region
/note="THE1B repeat: matches 1..364 of consensus"
23524..25074
repeat_region
/note="THE1B-INTERNAL repeat: matches 1..1580 of consensus"
25075..25447
repeat_region
/note="THE1B repeat: matches 1..364 of consensus"
complement(25519..25960)
misc_feature
/note="match: STS: Em:HSC685"
complement(26885..27263)
misc_feature
/note="match: GSS: Em:AQ877624"
27581..27822
repeat_region
/note="L1P412 repeat: matches -1422..-1192 of consensus"
27917..29879
repeat_region
/note="L1P413 repeat: matches 968..3008 of consensus"
30125..30409
repeat_region
/note="AluSc repeat: matches 1..285 of consensus"
30450..33568
repeat_region
/note="L1P413 repeat: matches 3003..6156 of consensus"
34046..34050
misc_feature
/note="137 bases of IS2 Transposon (V00610) removed here
This sequence represents the duplicated flanking sequence
of the IS2"
34565..35225
repeat_region
/note="L2 repeat: matches 1628..2327 of consensus"
35262..35440
repeat_region
/note="L2 repeat: matches 2515..2682 of consensus"
38611..38689
repeat_region
/note="MIR repeat: matches 176..249 of consensus"
38715..38881
misc_feature
/note="L1P413 repeat: matches -150..4 of consensus"
complement(38932..39495)
repeat_region
/note="match: STS: Em:HSC67C4"
39223..40001
repeat_region
/note="L1P413 repeat: matches -651..158 of consensus"
40235..41809
repeat_region
/note="L1P412 repeat: matches 522..2208 of consensus"
41810..42214
repeat_region
/note="MSTB repeat: matches 1..426 of consensus"
42224..43684
repeat_region
/note="MSTB-INTERNAL repeat: matches 198..1648 of consensus"
43685..44049
repeat_region
/note="THE1C repeat: matches 1..371 of consensus"
44050..44238
repeat_region
/note="MSTB-INTERNAL repeat: matches 9..198 of consensus"
44265..44651
repeat_region
/note="MSTB repeat: matches 1..408 of consensus"
44652..46249
repeat_region
/note="L1P repeat: matches 2204..3791 of consensus"
46252..46353
repeat_region
/note="L1P repeat: matches 3794..3897 of consensus"
46351..46955
repeat_region
/note="L1P repeat: matches 3816..4432 of consensus"
46924..47680
repeat_region
/note="L1P repeat: matches 3961..4718 of consensus"
47686..48489
repeat_region
/note="L1P47 repeat: matches 5302..6137 of consensus"
48518..50273
repeat_region
/note="L1P415 repeat: matches 4395..6157 of consensus"
50967..51212
repeat_region
/note="MIR repeat: matches 28..252 of consensus"
52867..52958
repeat_region
/note="46 copies 2 mer aa 67% conserved"
53142..53324
repeat_region
/note="MIR repeat: matches 6..189 of consensus"
53486..53618
repeat_region
/note="MIR repeat: matches 20..145 of consensus"
55100..55525
misc_feature
/note="match: GSS: Em:AQ828535"
55105..55472
misc_feature
/note="match: GSS: Em:AQ039042"
55327..55423
repeat_region
/note="MUT1J repeat: matches 400..494 of consensus"
56269..56320
repeat_region
/note="26 copies 2 mer ca 96% conserved"
57538..57957
repeat_region
/note="MER47B repeat: matches 1..418 of consensus"
57958..58214
repeat_region
/note="MUT1B repeat: matches 307..566 of consensus"
58395..58472
repeat_region
/note="MUT1D repeat: matches 30..101 of consensus"
59578..60160
repeat_region
/note="L1M1B repeat: matches 102..662 of consensus"
60270..61693
repeat_region
/note="L1M1B repeat: matches 673..2097 of consensus"
62055..62385
repeat_region
/note="MUT1C repeat: matches 139..466 of consensus"
62386..62431
repeat_region
/note="MUT1FB repeat: matches 366..411 of consensus"
62435..62504
repeat_region
/note="35 copies 2 mer tg 80% conserved"
62529..62910
Query Match 2.2%; Score 21; DB 9; Length 137154;
Best Local Similarity 100.0%; Pred.No.5.4;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 386 TATTTTGAATTCGATG 406
DB 11746 TATTTTGAATTCGATG 11766
```

```

RESULT 16
AC079562      140795 bp   DNA      linear   HTG 02-SEP-2000
LOCUS         Mus musculus clone RP23-465P16, WORKING DRAFT SEQUENCE. 24
DEFINITION   Unordered pieces.
ACCESSION    AC079562
VERSION      AC079562.1 GI:9964927
KEYWORDS     HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE       Mus musculus.
ORGANISM     Mus musculus.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE    1 (bases 1 to 140795)
AUTHORS      DOE Joint Genome Institute.
TITLE        Sequencing of Mouse
JOURNAL      Unpublished
REFERENCE    2 (bases 1 to 140795)
AUTHORS      DOE Joint Genome Institute.
TITLE        Direct Submission
JOURNAL      Submitted (02-SEP-2000) Production Sequencing Facility, DOE Joint
              Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT      -----Genome Center
              Center: Joint Genome Institute
              Center Code: JGI
              Web site: http://www.jgi.doe.gov
              -----
              Project Information
              Center Project Name: 1915808
              Center clone name: RPCT-23_465P16
              -----
              Summary Statistics
              Consensus quality: 112243 bases at least Q40
              Consensus quality: 123963 bases at least Q30
              Consensus quality: 126961 bases at least Q20
              Estimated insert size: 187000; agarose-ff estimation
              Estimated insert size: 138495; sum-of-ctdigs estimation
              Quality coverage: 6.23 in Q20 bases; agarose-ff estimation
              Quality coverage: 8.41 in Q20 bases; sum-of-ctdigs estimation.
              NOTE: This is a 'working draft' sequence. It currently
              * consists of 24 contigs. The true order of the pieces
              * is not known and their order in this sequence record is
              * arbitrary. Gaps between the contigs are represented as
              * runs of N, but the exact sizes of the gaps are unknown.
              * This record will be updated with the finished sequence
              * as soon as it is available and the accession number will
              * be preserved.
              -----
              1
              1202: contig of 1202 bp in length
              1203
              1302: gap of unknown length
              1303
              2446: contig of 1144 bp in length
              2447
              2546: gap of unknown length
              3844: contig of 1298 bp in length
              3845
              3944: gap of unknown length
              5015: contig of 1071 bp in length
              5016
              5115: gap of unknown length
              5116
              6372: contig of 1257 bp in length
              6472: gap of unknown length
              6473
              7613: contig of 1141 bp in length
              7713: gap of unknown length
              7714
              8895: contig of 1182 bp in length
              8896
              8995: gap of unknown length
              10604: contig of 1609 bp in length
              10605
              10704: gap of unknown length
              11824: contig of 1120 bp in length
              11825
              11924: gap of unknown length
              11925
              13682: contig of 1758 bp in length
              13683
              13782: gap of unknown length
              13783
              16337: contig of 2555 bp in length
              16338
              16437: gap of unknown length
              16438
              18700: contig of 2263 bp in length
              18701
              18800: gap of unknown length
              24180: contig of 5380 bp in length
              24181
              24280: gap of unknown length
              30054: contig of 5774 bp in length

```

```

* 30055 30154: gap of unknown length
* 30155 36771: contig of 6617 bp in length
* 36772 36871: gap of unknown length
* 36872 41167: contig of 4296 bp in length
* 41168 41267: gap of unknown length
* 41268 48106: contig of 6839 bp in length
* 48107 48206: gap of unknown length
* 48207 54870: contig of 6664 bp in length
* 54871 54970: gap of unknown length
* 54971 61950: contig of 6980 bp in length
* 61951 62050: gap of unknown length
* 62051 72051: contig of 10001 bp in length
* 72052 72151: gap of unknown length
* 72152 86891: contig of 14740 bp in length
* 86892 86991: gap of unknown length
* 86992 103983: contig of 16992 bp in length
* 103984 104083: gap of unknown length
* 104084 123755: contig of 19672 bp in length
* 123756 123855: gap of unknown length
* 123856 140795: contig of 16940 bp in length.
              Location/Qualifiers
              1. 140795
                 /organism="Mus musculus"
                 /db_xref="taxon:10090"
                 /clone="RP23-465P16"
                 /clone_lib="RPCT mouse BAC library 23"
BASE COUNT   33775 a 36314 c 35670 g 32605 t 2431 others
ORIGIN
Query Match      2.2%; Score 21; DB 2; Length 140795;
Best Local Similarity 100.0%; Pred. No. 5.4;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 725 CAAAGAAAGACAAAGACCAA 745
Db 16702 CAAAGAAAGACAAAGACCAA 16722
RESULT 17
LOCUS         AC010534      146385 bp   DNA      linear   PRI 26-JAN-2001
DEFINITION   Homo sapiens chromosome 16 clone RP11-259E8, complete sequence.
ACCESSION    AC010534
VERSION      AC010534.7 GI:12545287
KEYWORDS     HTG.
ORGANISM     Homo sapiens.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE    1 (bases 1 to 146385)
AUTHORS      DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE        Direct Submission
JOURNAL      Unpublished
REFERENCE    2 (bases 1 to 146385)
AUTHORS      DOE Joint Genome Institute.
TITLE        Direct Submission
JOURNAL      Submitted (15-SEP-1999) Production Sequencing Facility, DOE Joint
              Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
REFERENCE    3 (bases 1 to 146385)
AUTHORS      DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE        Direct Submission
JOURNAL      Submitted (26-JAN-2001) DOE Joint Genome Institute, 2800 Mitchell
              Drive, Walnut Creek, CA 94598, USA
COMMENT      On Jan 26, 2001 this sequence version replaced gi:9954624.
              Draft Sequence Produced by DOE Joint Genome Institute
              www.jgi.doe.gov
              Finishing Completed at Stanford Human Genome Center
              www-shgc.stanford.edu
              Quality: Phrap Quality >=40 99.8% of Sequence;
              Estimated Total Number of Errors is 0.2.
              STS Content:
              SHGC-80918 G53965.
              Location/Qualifiers

```

FEATURES

7721	7705	4925	5025	5927	6248
1912	1880	5249	5271	7454	7495
57	<800	1965	1977	3187	3202
2809	2850	802	741	1416	1424
188	<800	2883	2897	1489	1424
4011	4052	523	<800	696	<800
48	<800	746	741	9782	9681
763	765	857	872	4027	4017
1336	1342	1915	1977	68	<800
1486	1454	1481	1439	5698	5640
1136	1107	6467	6607	2913	2925
5365	5335	671	<800	1999	2007
2191	2190	5011	5025	810	822
3418	3454	400	<800	50	<800
111	<800	3989	3787	2317	2252
1032	1040	7133	7253	4222	4154
3078	3067	1323	1309	3052	3036
4342	4359	4730	4674	665	<800
111	<800	1226	1213	801	822
224	<800	948	975	170	<800
8194	8205	275	<800	2028	2007
2340	2387	3763	3787	1306	1317
7439	7705	10576	10547	6252	6248
1471	1454	3799	3787	5603	5543
153	<800	13397	13418	2039	2007
5613	5590	98	<800	188	<800
1379	1342	4810	4674	766	<800
4087	4052	200	<800	582	<800
1261	1243	1998	1977	1340	1317
539	<800	926	975	683	<800
4416	4359	3824	3787	9003	8843
		5578	5601	6390	6248
		1834	1794	543	<800
		1374	1309	5270	5161
		1862	1794	832	822
		1949	1977	5561	5543
		878	975	186	<800

Query Match 2.2%; Score 21; DB 9; Length 166496;
 Best Local Similarity 100.0%; Pred. No. 5.3;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 32 TATTGGAGAGGCATTAAATA 52

Db 24488 TATTGGAGAGGCATTAAATA 24468

RESULT 19

AC018846/c

LOCUS

DEFINITION

AC018846

AC018846.4

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Sequence Quality Assessment:

This entry has been annotated with sequence quality

estimates computed by the Phrap assembly program.

All manually edited bases have been reduced to quality zero.

Quality levels above 40 are expected to have less than

1 error in 10,000 bp.

Base-by-base quality values are not generally visible from the

Genbank flat file format but are available as part

of this entry's ASN.1 file.

Sequence Quality Assessment:

This entry has been annotated with sequence quality

estimates computed by the Phrap assembly program.

All manually edited bases have been reduced to quality zero.

Quality levels above 40 are expected to have less than

1 error in 10,000 bp.

Base-by-base quality values are not generally visible from the

Genbank flat file format but are available as part

of this entry's ASN.1 file.

FEATURES

source

1. 175518

/organism="Homo sapiens"

/db_xref="taxon:9606"

/chromosome="16"

/clone="RP11-370P15"

BASE COUNT

57156 a 32486 c 31996 g 53880 t

ORIGIN

Query Match 2.2%; Score 21; DB 9; Length 175518;

Best Local Similarity 100.0%; Pred. No. 5.3;

Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;


```
misc_feature /note="assembly_name:Contig72"
26495..31534
/note="assembly_name:Contig73"
misc_feature 31635..37164
/note="assembly_name:Contig74"
misc_feature 37265..44921
/note="assembly_name:Contig75
clone_end:SP6
vector_side:right"
misc_feature 45022..50595
/note="assembly_name:Contig76"
50696..59103
/note="assembly_name:Contig77"
misc_feature 59204..69267
/note="assembly_name:Contig78"
misc_feature 69368..82673
/note="assembly_name:Contig79"
misc_feature 82774..97615
/note="assembly_name:Contig80"
97716..115178
/note="assembly_name:Contig81"
misc_feature 115279..128781
/note="assembly_name:Contig82"
misc_feature 128882..144470
/note="assembly_name:Contig83"
misc_feature 144571..161750
/note="assembly_name:Contig84"
misc_feature 161851..187038
/note="assembly_name:Contig85"
BASE COUNT 55977 a 36931 c 36406 g 55111 t 2613 others
ORIGIN
```

```
Query Match 2.2%; Score 21; DB 2; Length 187038;
Best Local Similarity 100.0%; Pred. No. 5.3;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 17 TTGAAGGACAAATATTGG 37
|||||
Db 37096 TTGAAGGACAAATATTGG 37076
```

```
RESULT 21
AC019228 196216 bp DNA linear HTG 07-JUL-2000
LOCUS Homo sapiens chromosome 1L clone RP11-570N12, WORKING DRAFT
DEFINITION
AC019228 SEQUENCE, 20 unordered pieces.
ACCESSION AC019228.5 GI:8568940
VERSION HTG; HTGS_PHASE1; HTGS_DRAFT.
KEYWORDS
SOURCE Homo sapiens.
ORGANISM Homo sapiens.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 196216)
AUTHORS Waterston,R.H.
TITLE The sequence of Homo sapiens clone
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 196216)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (30-DEC-1999) Genome Sequencing Center, Washington
MO 63108, USA
UNIVERSITY School of Medicine, 4444 Forest Park Parkway, St. Louis,
On Jun 16, 2000 this sequence version replaced gi:7233983.
```

```
----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
----- Project Information -----
Center project name: H.NH0570N12
----- Summary Statistics -----
Sequencing vector: M13; 67%
```

Sequencing vector: plasmid; 33%
Chemistry: Dye-primer ET; 67% of reads
Chemistry: Dye-terminator Big Dye; 33% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 182091 bases at least Q40
Consensus quality: 186866 bases at least Q30
Consensus quality: 189092 bases at least Q20
Insert size: 185000; agarose-fp
Insert size: 194316; sum-of-contigs
Quality coverage: 3.88 in Q20 bases; agarose-fp
Quality coverage: 3.68 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 20 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1	1722: contig of 1722 bp in length
*	1723
*	1822: gap of unknown length
*	1823
*	3457: contig of 1635 bp in length
*	3458
*	3557: gap of unknown length
*	3558
*	6192: contig of 2635 bp in length
*	6193
*	6292: gap of unknown length
*	6293
*	9326: contig of 3034 bp in length
*	9327
*	9426: gap of unknown length
*	9427
*	12746: contig of 3320 bp in length
*	12747
*	12846: gap of unknown length
*	12847
*	16489: contig of 3643 bp in length
*	16490
*	16589: gap of unknown length
*	16590
*	19714: contig of 3125 bp in length
*	19715
*	19814: gap of unknown length
*	19815
*	26224: contig of 6410 bp in length
*	26225
*	26324: gap of unknown length
*	26325
*	31243: contig of 4919 bp in length
*	31244
*	31344: gap of unknown length
*	31344
*	38835: contig of 7492 bp in length
*	38836
*	38935: gap of unknown length
*	38936
*	46701: contig of 7766 bp in length
*	46702
*	46801: gap of unknown length
*	46802
*	57002: contig of 10201 bp in length
*	57003
*	57102: gap of unknown length
*	57103
*	66636: contig of 9534 bp in length
*	66637
*	66736: gap of unknown length
*	66737
*	76277: contig of 9341 bp in length
*	76278
*	76377: gap of unknown length
*	76378
*	87128: contig of 10751 bp in length
*	87129
*	87228: gap of unknown length
*	87229
*	98549: contig of 11321 bp in length
*	98550
*	98650
*	109482: gap of unknown length
*	109493
*	109592: contig of 10843 bp in length
*	109593
*	124679: gap of unknown length
*	124680
*	124779: contig of 15087 bp in length
*	124780
*	155821: contig of 31042 bp in length
*	155822
*	155922
*	155922
*	196216: contig of 40285 bp in length.

FEATURES

source

```
1..196216
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="1L"
/clone="RP11-570N12"
1..1722
/note="assembly_name:Contig8"
1823..3457
/note="assembly_name:Contig10"
3558..6192
/note="assembly_name:Contig11
clone_end:SP6
vector_side:right"
```

```

misc_feature      6293..9326
                  /note="assembly_name:Contig12"
misc_feature      9427..12746
                  /note="assembly_name:Contig13"
misc_feature      12847..16489
                  /note="assembly_name:Contig14"
misc_feature      16590..19714
                  /note="assembly_name:Contig15
clone_end::T7
vector_side:left"
misc_feature      19815..26224
                  /note="assembly_name:Contig16"
misc_feature      26325..31243
                  /note="assembly_name:Contig17"
misc_feature      31344..38835
                  /note="assembly_name:Contig18"
misc_feature      38936..46701
                  /note="assembly_name:Contig19"
misc_feature      46802..57002
                  /note="assembly_name:Contig20"
misc_feature      57103..66636
                  /note="assembly_name:Contig21"
misc_feature      66737..76277
                  /note="assembly_name:Contig22"
misc_feature      76378..87128
                  /note="assembly_name:Contig23"
misc_feature      87229..98549
                  /note="assembly_name:Contig24"
misc_feature      98650..109492
                  /note="assembly_name:Contig25"
misc_feature      109593..124676
                  /note="assembly_name:Contig26"
misc_feature      124780..155821
                  /note="assembly_name:Contig27"
misc_feature      155922..196216
                  /note="assembly_name:Contig28"
BASE COUNT      57961 a 35701 c 36544 g 64086 t 1924 others
ORIGIN
Query Match      2.2%; Score 21; DB 2; Length 196216;
Best Local Similarity 100.0%; Pred. No. 5.3;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 386 TATTTTGAATTCGATG 406
|||||
Db 63025 TATTTTGAATTCGATG 63045

```

RESULT 22
AC084217 216988 bp DNA linear ROD 15-MAY-2002
LOCUS Mus Musculus Strain C57BL/6J Chromosome 5 RP23-327P24, complete
DEFINITION
AC084217 GI:13549250
VERSION
KEYWORDS
SOURCE
ORGANISM
Mus musculus.
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 216988)
Montgomery,K.T., Grills,G., Han,J., Chiu,D., Decker,J., Fusina,M.,
Goltz,J., Halder,A., Hall,L., Ioshikhes,I.P., Lee,E., Long,J.,
Pereira,A., Shim,C., Thomas,E. and Kucherlapati,R.
High Throughput Mouse Sequencing
Unpublished
2 (bases 1 to 216988)
Montgomery,K.T., Grills,G., Han,J., Chiu,D., Decker,J., Fusina,M.,
Goltz,J., Halder,A., Hall,L., Ioshikhes,I.P., Lee,E., Long,J.,
Pereira,A., Shim,C., Thomas,E. and Kucherlapati,R.
Direct Submission
Submitted (17-OCT-2000) Department of Molecular Genetics, Albert
Einstein College of Medicine Genome Center, 1300 Morris Park Ave.,

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
Bronx, NY 10461, USA
3 (bases 1 to 216988)
Montgomery,K.T., Grills,G., Han,J., Chiu,D., Decker,J., Fusina,M.,
Goltz,J., Halder,A., Hall,L., Ioshikhes,I.P., Lee,E., Long,J.,
Pereira,A., Shim,C., Thomas,E. and Kucherlapati,R.
Direct Submission
Submitted (05-APR-2001) Department of Molecular Genetics, Albert
Einstein College of Medicine Genome Center, 1300 Morris Park Ave.,
Bronx, NY 10461, USA
On Apr 5, 2001 this sequence version replaced gi:11177930.
-----Genome Center:
Harvard Partners Genome Center
Center Code: HPGC
Web site: <http://www.hpgc.org/Sequence/mouse.html>
Contact: hpgc@mednet.mgh.harvard.edu

CLONE LENGTH: This sequence represents the entire insert of this
clone unless otherwise noted. If there are overlapping clones, the
overlaps are noted in the beginning and end of the Features
listing.

ANNOTATION OF FEATURES:

STSs are identified using ePCR (Genome Res. 7:541-550).
Repeats are identified using RepeatMasker (A. Smit and P. Green,
unpublished.) for Human and Mouse sequences.
Genes and Regions of sequence similarity are identified by BLAST
(Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to EST
and cDNA sequences in Unigene. Genes demonstrate at least two exons
flanked by consensus splice sites that maintain sequence continuity
across the splice junctions. Sequences that are not identical
matches are annotated as similar.

SEQUENCING READ COVERAGE: Attempts are made to complete double
stranded sequence for all regions. All sequence is completed to a
standard of coverage with a minimum of 3 reads with no ambiguities.
If the sequence coverage for a region does not meet this standard,
it is indicated in the annotation as Low Coverage. Low coverage
linkages are verified by PCR product size verification or
verification of forward and reverse reads from clones which span
the low coverage area.

QUALITY OF INDIVIDUAL BASES: This sequence meets stringent quality
standards - estimated average error rate is less than 1 per 10,000
bases using the Consed quality parameters. Regions that do not
meet this requirement are annotated as Low Quality.

-----Summary Statistics

Center project name: ACP
Sequencing vector: pUC18; L08752
Chemistry: Dye-terminator Big Dye; 100%
Assembly program: Phrap version 0.990319
Contig length: 216988
Fraction of Phrap value < 40: 0.00216
Error Rate in Consed: 0.02 per 10,000 bases
Number of N's in consensus: 0

----- Distribution of Quality < 40 Bases:

10001									
9001									
8001									
7001									
6001									
5001									
4001									
3001									
2001									
1001									
01		*	*	*	*	*	*	*	*
5	10	15	20	25	30	35	40		
Phrap Value Range									

Sequencing vector: plasmid; 100%
 Chemistry: Dye-primer ET; 0% of reads
 Assembly: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; version 0.990319
 Consensus quality: 239545 bases at least Q40
 Consensus quality: 240084 bases at least Q40
 Consensus quality: 240487 bases at least Q20
 Insert size: 242781; sum-of-contigs
 Quality coverage: 15.02 in Q20 bases; agarose-fp
 Quality coverage: 10.67 in Q20 bases; sum-of-contigs

NOTE: This is a 'working draft' sequence. It currently consists of 7 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

1 2418: contig of 2418 bp in length
 * 2419 2518: gap of unknown length
 * 2519 5398: contig of 2880 bp in length
 * 5399 5498: gap of unknown length
 * 5499 30586: contig of 25088 bp in length
 * 30587 30686: gap of unknown length
 * 30687 105664: contig of 74978 bp in length
 * 105665 105764: gap of unknown length
 * 105765 144553: contig of 38689 bp in length
 * 144554 144554: gap of unknown length
 * 144554 240735: contig of 96182 bp in length
 * 240736 240835: gap of unknown length
 * 240836 243381: contig of 2546 bp in length.

FEATURES
 source
 1. .243381
 /organism="Mus musculus"
 /db_xref="taxon:10090"
 /chromosome="UNK"
 /clone="Rp23-23812"

misc-feature
 1. .2418
 /note="assembly_name:Contig10"
 2519. .5398
 /note="assembly_name:Contig11"
 5499. .30586
 /note="assembly_name:Contig12"
 30687. .105664
 /note="assembly_name:Contig14"
 105765. .144553
 /note="assembly_name:Contig13"
 144554. .240735
 /note="assembly_name:Contig15"
 240836. .243381
 /note="assembly_name:Contig19"
 misc-feature
 65599 a 55671 c 55170 g 66312 t 629 others

Query Match 2.2%: Score 21; DB 2; Length 243381;
 Best Local Similarity 100.0%: Pred. No. 5.2;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 655 GGAAGAAAGTGGGATCTCA 675
 ||||||||||||||||||
 Db 214109 GGAAGAAAGTGGGATCTCA 214129

RESULT 24
 E06193/c 849 bp RNA linear PAT 29-SEP-1997
 LOCUS
 DEFINITION CDNA encoding genes derived from hepatitis C virus.
 ACCESSION E06193
 VERSION E06193.1 GI:2174380
 KEYWORDS JP 1994000085-A/33.
 SOURCE Hepatitis C virus.

ORGANISM
 Hepatitis C virus
 Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 Hepacivirus.
 1 (bases 1 to 849)
 AUTHORS Seki,M., Honda,Y., Takahashi,K., Murakami,T., Teranishi,Y. and Hayashi,N.
 TITLE GENE OR DNA FRAGMENT DERIVED FROM HEPATITIS C VIRUS, POLYPEPTIDE CODED BY THE SAME AND ITS PRODUCTION
 JOURNAL Patent: JP 1994000085-A 33 11-JAN-1994;
 MITSUBISHI KASEI CORP
 OS (hepatitis C virus)
 PN JP 1994000085-A/33
 PD 11-JAN-1994
 PE 11-JUN-1992 JP 1992194497
 PR 11-JUN-1991 JP 91P 139268, 12-JUL-1991 JP 91P 172794, PR 07-OCT-1991 JP 91P 287008, 16-DEC-1991 JP 91P 332329, PR 20-APR-1992 JP 92P 99957
 PI SEKI MAKOTO, HONDA YOSHIKAZU, TAKAHASHI KAZUNOBU, PI MURAKAMI TOMOKO,
 PI TERANISHI YUTAKA, HAYASHI NORIO
 PC C12N15/51,C07K7/06,C07K7/08,C07K7/10,C07K13/00,C07K15/12, PC C12N1/21,C12N5/10,
 PC C12N15/11,C12N15/70,C12N15/85,C12P21/02//A61K39/00,A61K39/29, PC (C12N1/21),
 PC C12R1:19),(C12N5/10,C12R1:91),(C12P21/02,C12R1:19), (C12P21/02, PC C12R1:91),
 PC C07K99:00;
 CC strandedness: Double;
 CC topology: linear;
 CC anti-sense: No;
 CC key *source: clone-MX25;
 FH Location/Qualifiers
 FT mat-peptide 1. .849
 FT /product='the peptides reacting specifically and immunochemically with the serum of hepatitis type C patient';

FEATURES
 source
 1. .849
 /organism="Hepatitis C virus"
 /db_xref="taxon:11103"

BASE COUNT 132 a 227 c 232 g 188 t 70 others

Query Match 2.1%: Score 20; DB 6; Length 849;
 Best Local Similarity 100.0%: Pred. No. 25;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 311 CAGAGACACAGAGAGAGA 330
 ||||||||||||||||||
 Db 125 CAGAGACACAGAGAGAGA 106

RESULT 25
 E06204/c 849 bp RNA linear PAT 29-SEP-1997
 LOCUS
 DEFINITION CDNA encoding genes derived from hepatitis C virus.
 ACCESSION E06204
 VERSION E06204.1 GI:2174391
 KEYWORDS JP 1994000085-A/44.
 SOURCE Hepatitis C virus.
 ORGANISM Hepatitis C virus
 Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 Hepacivirus.
 1 (bases 1 to 849)

REFERENCE
 AUTHORS Seki,M., Honda,Y., Takahashi,K., Murakami,T., Teranishi,Y. and Hayashi,N.
 TITLE GENE OR DNA FRAGMENT DERIVED FROM HEPATITIS C VIRUS, POLYPEPTIDE CODED BY THE SAME AND ITS PRODUCTION
 JOURNAL Patent: JP 1994000085-A 44 11-JAN-1994;
 MITSUBISHI KASEI CORP

COMMENT OS (hepatitis C virus)
PN JP 1994000085-A/44
PD 11-JAN-1994
PR 11-JUN-1992 JP 1992194497
PR 11-JUN-1991 JP 91P 139268, 12-JUL-1991 JP 91P 172794, PR
07-OCT-1991 JP 91P 287008, 16-DEC-1991 JP 91P 332329, PR
20-APR-1992 JP 92P 99957
PI SEKI MAKOTO, HONDA YOSHIKAZU, TAKAHASHI KAZUNOBU, PI
MURAKAMI TOMOKO,
PI TERANISHI YUTAKA, HAYASHI NORIO
PC C12N1/21,C12N5/10,C12N15/70,C12N15/85,C12P21/02//A61K39/00,A61K39/29,
C12N1/21,C12N5/10,
PC C12N15/11,C12N15/70,C12N15/85,C12P21/02//A61K39/00,A61K39/29,
PC (C12N1/21,
PC C12R1:19),(C12N5/10,C12R1:91),(C12P21/02,C12R1:19),(C12P21/02,
PC C12R1:91),
PC C07K99:00;
CC strandedness: Double;
CC topology: Linear;
CC anti-sense: No;
CC *source: clone-MX25-1;
FH Key Location/Qualifiers
FT mat_peptide 1..849
FT /product='the peptides reacting specifically
FT and
FT immunochemically with the serum of hepatitis
FT type C
FT patient'.
FT Location/Qualifiers
FEATURES
source 1..849
/organism="Hepatitis C virus"
/db_xref="taxon:11103"
BASE COUNT 146 a 244 c 249 g 210 t
ORIGIN
Query Match 2.1%; Score 20; DB 6; Length 849;
Best Local Similarity 100.0%; Pred. No. 25;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 311 CAGAAGACACAGGAAGA 330
Db 125 CAGAAGACACAGGAAGA 106
RESULT 26
E06205/C 849 bp RNA linear PAT 29-SEP-1997
LOCUS E06205
DEFINITION cDNA encoding genes derived from hepatitis C virus.
ACCESSION E06205.1 GI:2174392
VERSION JP 1994000085-A/45.
KEYWORDS Hepatitis C virus.
SOURCE Hepatitis C virus.
ORGANISM Hepatitis C virus.
VIRUSES; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
Hepacivirus.
1 (bases 1 to 849)
SEKI,M., Honda,Y., Takahashi,K., Murakami,T., Teranishi,Y. and
Hayashi,N.
GENE OR DNA FRAGMENT DERIVED FROM HEPATITIS C VIRUS, POLYPEPTIDE
CODED BY THE SAME AND ITS PRODUCTION
PATENT: JP 1994000085-A 45 11-JAN-1994;
JOURNAL MITSUBISHI KASEI CORP
OS (hepatitis C virus)
PN JP 1994000085-A/45
PD 11-JAN-1994
PR 11-JUN-1992 JP 1992194497
PR 11-JUN-1991 JP 91P 139268, 12-JUL-1991 JP 91P 172794, PR
07-OCT-1991 JP 91P 287008, 16-DEC-1991 JP 91P 332329, PR
20-APR-1992 JP 92P 99957
PI SEKI MAKOTO, HONDA YOSHIKAZU, TAKAHASHI KAZUNOBU, PI
MURAKAMI TOMOKO,
PI TERANISHI YUTAKA, HAYASHI NORIO

PC C12N15/51,C07K7/06,C07K7/08,C07K7/10,C07K13/00,C07K15/12, PC
C12N1/21,C12N5/10,
PC C12N15/11,C12N15/70,C12N15/85,C12P21/02//A61K39/00,A61K39/29,
PC (C12N1/21,
PC C12R1:19),(C12N5/10,C12R1:91),(C12P21/02,C12R1:19),(C12P21/02,
PC C12R1:91),
PC C07K99:00;
CC strandedness: Double;
CC topology: Linear;
CC anti-sense: No;
CC *source: clone-MX25-2;
FH Key Location/Qualifiers
FT mat_peptide 1..849
FT /product='the peptides reacting specifically
FT and
FT immunochemically with the serum of hepatitis
FT type C
FT patient'.
FT Location/Qualifiers
FEATURES
source 1..849
/organism="Hepatitis C virus"
/db_xref="taxon:11103"
BASE COUNT 150 a 251 c 246 g 202 t
ORIGIN
Query Match 2.1%; Score 20; DB 6; Length 849;
Best Local Similarity 100.0%; Pred. No. 25;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 311 CAGAAGACACAGGAAGA 330
Db 125 CAGAAGACACAGGAAGA 106
RESULT 27
E06206/C 849 bp RNA linear PAT 29-SEP-1997
LOCUS E06206
DEFINITION cDNA encoding genes derived from hepatitis C virus.
ACCESSION E06206
VERSION E06206.1 GI:2174393
KEYWORDS JP 1994000085-A/46.
SOURCE Hepatitis C virus.
ORGANISM Hepatitis C virus.
VIRUSES; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
Hepacivirus.
1 (bases 1 to 849)
SEKI,M., Honda,Y., Takahashi,K., Murakami,T., Teranishi,Y. and
Hayashi,N.
GENE OR DNA FRAGMENT DERIVED FROM HEPATITIS C VIRUS, POLYPEPTIDE
CODED BY THE SAME AND ITS PRODUCTION
PATENT: JP 1994000085-A 46 11-JAN-1994;
JOURNAL MITSUBISHI KASEI CORP
OS (hepatitis C virus)
PN JP 1994000085-A/46
PD 11-JAN-1994
PR 11-JUN-1992 JP 1992194497
PR 11-JUN-1991 JP 91P 139268, 12-JUL-1991 JP 91P 172794, PR
07-OCT-1991 JP 91P 287008, 16-DEC-1991 JP 91P 332329, PR
20-APR-1992 JP 92P 99957
PI SEKI MAKOTO, HONDA YOSHIKAZU, TAKAHASHI KAZUNOBU, PI
MURAKAMI TOMOKO,
PI TERANISHI YUTAKA, HAYASHI NORIO
PC C12N1/21,C12N5/10,C12N15/70,C12N15/85,C12P21/02//A61K39/00,A61K39/29,
C12N1/21,C12N5/10,
PC C12N15/11,C12N15/70,C12N15/85,C12P21/02//A61K39/00,A61K39/29,
PC (C12N1/21,
PC C12R1:19),(C12N5/10,C12R1:91),(C12P21/02,C12R1:19),(C12P21/02,
PC C12R1:91),
PC C07K99:00;
CC strandedness: Double;
CC topology: Linear;
CC anti-sense: No;

CC *source: clone-MX25-3;
FH Key Location/Qualifiers
FT mat-peptide 1..849
FT /product='the peptides reacting specifically
FT and
FT immunochemically with the serum of hepatitis
FT type C
FT patient'
FT Location/Qualifiers
FEATURES
Source 1..849
/organism='Hepatitis C virus'
/db_xref='taxon:11103'
BASE COUNT 144 a 246 c 248 g 211 t
ORIGIN

Query Match 2.1%; Score 20; DB 6; Length 849;
Best Local Similarity 100.0%; Pred. No. 25;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 311 CAGAGAACACAGGAAGA 330
|||||
Db 125 CAGAGAACACAGGAAGA 106

RESULT 28
E06389 849 bp RNA linear PAT 29-SEP-1997
LOCUS E06389
DEFINITION cDNA encoding genes derived from hepatitis C virus.
ACCESSION E06389
VERSION E06389.1 GI:2174576
KEYWORDS JP 1994000086-A/33.
SOURCE Hepatitis C virus.
ORGANISM Hepatitis C virus.
Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
Hepacivirus.
1 (bases 1 to 849)
Seki,M., Honda,Y., Takahashi,K., Murakami,T., Teranishi,Y. and
Hayashi,N.
GENE FOR DNA FRAGMENT DERIVED FROM HEPATITIS C VIRUS, POLYPEPTIDE
CODED BY THE SAME AND ITS PRODUCTION
Patent: JP 1994000086-A 33 11-JAN-1994;
MITSUBISHI KASEI CORP
OS (hepatitis C virus)
PN JP 1994000086-A/33
PD 11-JAN-1994
PF 07-OCT-1992 JP 1992293734
PR 07-OCT-1991 JP 91P 287008, 16-DEC-1991 JP 91P 332329, PR
20-APR-1992 JP 92P 99957
PI SEKI MAKOTO, HONDA YOSHIKAZU, TAKAHASHI KAZUNOBU, PI
MURAKAMI TOMOKO,
PI TERANISHI YUTAKA, HAYASHI NORIO
PC C12N15/51,C07K7/06,C07K7/10,C07K13/00,C12N5/10, PC
C12N15/11,
PC C12N15/85,C12P21/02//A61K39/29,(C12P21/02,C12R1:91),C07K99:00;
CC strandedness: Double;
CC topology: Linear;
CC anti-sense: No;
CC *source: clone-MX25-1;
FH Key Location/Qualifiers
FH mat-peptide 1..849
FH /product='the peptides reacting specifically
FH and
FH immunochemically with the serum of hepatitis
FH type C
FH patient'
FH Location/Qualifiers
FEATURES
Source 1..849
/organism='Hepatitis C virus'
/db_xref='taxon:11103'
BASE COUNT 132 a 227 c 232 g 188 t 70 others
ORIGIN

Query Match 2.1%; Score 20; DB 6; Length 849;
Best Local Similarity 100.0%; Pred. No. 25;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 311 CAGAGAACACAGGAAGA 330
|||||
Db 125 CAGAGAACACAGGAAGA 106

RESULT 29
E06400 849 bp RNA linear PAT 29-SEP-1997
LOCUS E06400
DEFINITION cDNA encoding genes derived from hepatitis C virus.
ACCESSION E06400
VERSION E06400.1 GI:2174587
KEYWORDS JP 1994000086-A/44.
SOURCE Hepatitis C virus.
ORGANISM Hepatitis C virus.
Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
Hepacivirus.
1 (bases 1 to 849)
Seki,M., Honda,Y., Takahashi,K., Murakami,T., Teranishi,Y. and
Hayashi,N.
GENE FOR DNA FRAGMENT DERIVED FROM HEPATITIS C VIRUS, POLYPEPTIDE
CODED BY THE SAME AND ITS PRODUCTION
Patent: JP 1994000086-A 44 11-JAN-1994;
MITSUBISHI KASEI CORP
OS (hepatitis C virus)
PN JP 1994000086-A/44
PD 11-JAN-1994
PF 07-OCT-1992 JP 1992293734
PR 07-OCT-1991 JP 91P 287008, 16-DEC-1991 JP 91P 332329, PR
20-APR-1992 JP 92P 99957
PI SEKI MAKOTO, HONDA YOSHIKAZU, TAKAHASHI KAZUNOBU, PI
MURAKAMI TOMOKO,
PI TERANISHI YUTAKA, HAYASHI NORIO
PC C12N15/51,C07K7/06,C07K7/10,C07K13/00,C12N5/10, PC
C12N15/11,
PC C12N15/85,C12P21/02//A61K39/29,(C12P21/02,C12R1:91),C07K99:00;
CC strandedness: Double;
CC topology: Linear;
CC anti-sense: No;
CC *source: clone-MX25-1;
FH Key Location/Qualifiers
FH mat-peptide 1..849
FH /product='the peptides reacting specifically
FH and
FH immunochemically with the serum of hepatitis
FH type C
FH patient'
FH Location/Qualifiers
FEATURES
Source 1..849
/organism='Hepatitis C virus'
/db_xref='taxon:11103'
BASE COUNT 146 a 244 c 249 g 210 t
ORIGIN

Query Match 2.1%; Score 20; DB 6; Length 849;
Best Local Similarity 100.0%; Pred. No. 25;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 311 CAGAGAACACAGGAAGA 330
|||||
Db 125 CAGAGAACACAGGAAGA 106

RESULT 30
E06401 849 bp RNA linear PAT 29-SEP-1997
LOCUS E06401
DEFINITION cDNA encoding genes derived from hepatitis C virus.
ACCESSION E06401

VERSION E06401.1 GI:2174588
KEYWORDS JP 1994000086-A/45.
SOURCE Hepatitis C virus.
ORGANISM Hepatitis C virus
Viruses: ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
Hepacivirus.
1 (bases 1 to 849)
REFERENCE 1 Seki,M., Honda,Y., Takahashi,K., Murakami,T., Teranishi,Y. and Hayashi,N.
TITLE GENE FOR DNA FRAGMENT DERIVED FROM HEPATITIS C VIRUS, POLYPEPTIDE CODED BY THE SAME AND ITS PRODUCTION
JOURNAL Patent: JP 1994000086-A 45 11-JAN-1994;
MITSUBISHI KASEI CORP
COMMENT OS (hepatitis C virus)
PN JP 1994000086-A/45
PD 11-JAN-1994
PR 07-OCT-1992 JP 1992293734
PR 07-OCT-1991 JP 91P 287008, 16-DEC-1991 JP 91P 332329, PR 20-APR-1992 JP 92P 99957
PI SEKI MAKOTO, HONDA YOSHIKAZU, TAKAHASHI KAZUNOBU, PI MURAKAMI TOMOKO,
PI TERANISHI YUTAKA, HAYASHI NORIO
PC C12N15/51,C07K7/06,C07K7/10,C07K13/00,C12N5/10, PC C12N15/11,
PC C12N15/85,C12P21/02//A61K39/29,(C12P21/02,C12R1:91),C07K99:00;
CC strandedness: Double;
CC topology: Linear;
CC anti-sense: No;
CC *source: clone-MX25-2;
FH key Location/Qualifiers
FT mat_peptide 1..849
FT /product='the peptides reacting specifically and immunochemically with the serum of hepatitis type C patient'.
FT Location/Qualifiers
FEATURES
source 1..849
location/Qualifiers
BASE COUNT 150 a 251 c 246 g 202 t
ORIGIN
Query Match 2.1%; Score 20; DB 6; Length 849;
Best Local Similarity 100.0%; Pred. No. 25;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 311 CAGAGAACACAGAGAAGA 330
IIIIIIIIIIIIIIIIIIII
Db 125 CAGAGAACACAGAGAAGA 106
RESULT 31
E06402/c 849 bp RNA linear PAT 29-SEP-1997
LOCUS
DEFINITION CDNA encoding genes derived from hepatitis C virus.
ACCESSION E06402
VERSION E06402.1 GI:2174589.
KEYWORDS JP 1994000086-A/46.
SOURCE Hepatitis C virus.
ORGANISM Hepatitis C virus
Viruses: ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
Hepacivirus.
1 (bases 1 to 849)
REFERENCE 1 Seki,M., Honda,Y., Takahashi,K., Murakami,T., Teranishi,Y. and Hayashi,N.
TITLE GENE FOR DNA FRAGMENT DERIVED FROM HEPATITIS C VIRUS, POLYPEPTIDE CODED BY THE SAME AND ITS PRODUCTION
JOURNAL Patent: JP 1994000086-A 46 11-JAN-1994;
MITSUBISHI KASEI CORP
COMMENT OS (hepatitis C virus)
PN JP 1994000086-A/46

PD 11-JAN-1994
PR 07-OCT-1992 JP 1992293734
PR 07-OCT-1991 JP 91P 287008, 16-DEC-1991 JP 91P 332329, PR 20-APR-1992 JP 92P 99957
PI SEKI MAKOTO, HONDA YOSHIKAZU, TAKAHASHI KAZUNOBU, PI MURAKAMI TOMOKO,
PI TERANISHI YUTAKA, HAYASHI NORIO
PC C12N15/51,C07K7/06,C07K7/10,C07K13/00,C12N5/10, PC C12N15/11,
PC C12N15/85,C12P21/02//A61K39/29,(C12P21/02,C12R1:91),C07K99:00;
CC strandedness: Double;
CC topology: Linear;
CC anti-sense: No;
CC *source: clone-MX25-3;
FH key Location/Qualifiers
FT mat_peptide 1..849
FT /product='the peptides reacting specifically and immunochemically with the serum of hepatitis type C patient'.
FT Location/Qualifiers
FEATURES
source 1..849
location/Qualifiers
BASE COUNT 144 a 246 c 248 g 211 t
ORIGIN
Query Match 2.1%; Score 20; DB 6; Length 849;
Best Local Similarity 100.0%; Pred. No. 25;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 311 CAGAGAACACAGAGAAGA 330
IIIIIIIIIIIIIIIIIIII
Db 125 CAGAGAACACAGAGAAGA 106
RESULT 32
E06200/c 1280 bp RNA linear PAT 29-SEP-1997
LOCUS
DEFINITION CDNA encoding genes derived from hepatitis C virus.
ACCESSION E06200
VERSION E06200.1 GI:2174387
KEYWORDS JP 1994000085-A/40.
SOURCE Hepatitis C virus.
ORGANISM Hepatitis C virus
Viruses: ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
Hepacivirus.
1 (bases 1 to 1280)
REFERENCE 1 Seki,M., Honda,Y., Takahashi,K., Murakami,T., Teranishi,Y. and Hayashi,N.
TITLE GENE OR DNA FRAGMENT DERIVED FROM HEPATITIS C VIRUS, POLYPEPTIDE CODED BY THE SAME AND ITS PRODUCTION
JOURNAL Patent: JP 1994000085-A 40 11-JAN-1994;
MITSUBISHI KASEI CORP
COMMENT OS (hepatitis C virus)
PN JP 1994000085-A/40
PD 11-JAN-1994
PR 11-JUN-1992 JP 1992194497
PR 11-JUN-1991 JP 91P 139268, 12-JUL-1991 JP 91P 172794, PR 07-OCT-1991 JP 91P 287008, 16-DEC-1991 JP 91P 332329, PR 20-APR-1992 JP 92P 99957
PI SEKI MAKOTO, HONDA YOSHIKAZU, TAKAHASHI KAZUNOBU, PI MURAKAMI TOMOKO,
PI TERANISHI YUTAKA, HAYASHI NORIO
PC C12N15/51,C07K7/06,C07K7/10,C07K13/00,C07K15/12, PC C12N1/21,C12N5/10,
PC C12N15/11,C12N15/70,C12N15/85,C12P21/02//A61K39/29,
PC C12N1/21,
PC C12R1:19,(C12N5/10,C12R1:91),(C12P21/02,C12R1:19),(C12P21/02, PC C07K99:00);

CC strandedness: Double;
CC topology: Linear;
CC anti-sense: No;
CC *source: clone-MX25026;
PH key Location/Qualifiers
FT mat_peptide 1..1278
FT /product='the peptides reacting specifically
FT and
FT immunochemically with the serum of hepatitis
FT type C
FT patient',
FT Location/Qualifiers
FEATURES
source Location/Qualifiers
1..1280
/organism="Hepatitis C virus"
/db_xref="taxon:11103"
BASE COUNT 203 a 354 c 356 g 269 t 98 others
ORIGIN
Query Match 2.1%; Score 20; DB 6; Length 1280;
Best Local Similarity 100.0%; Pred. No. 24;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 311 CAGAGAACACAGAGAGA 330
|||||
Db 125 CAGAGAACACAGAGAGA 106
RESULT 33
E06216/c 1280 bp RNA linear PAT 29-SEP-1997
LOCUS
DEFINITION cDNA encoding genes derived from hepatitis C virus.
ACCESSION E06216
VERSION E06216.1 GI:2174403
KEYWORDS JP 1994000085-A/56.
SOURCE Hepatitis C virus.
ORGANISM Hepatitis C virus
Virus; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
Hepacivirus.
REFERENCE 1 (bases 1 to 1280)
AUTHORS Seki,M., Honda,Y., Takahashi,K., Murakami,T., Teranishi,Y. and Hayashi,N.
TITLE GENE OR DNA FRAGMENT DERIVED FROM HEPATITIS C VIRUS, POLYPEPTIDE CODED BY THE SAME AND ITS PRODUCTION
JOURNAL Patent: JP 1994000085-A 56 11-JAN-1994;
MITSUBISHI KASEI CORP
COMMENT OS (hepatitis C virus)
PN JP 1994000085-A/56
PD 11-JAN-1994
PE 11-JUN-1992 JP 1992194497
PR 11-JUN-1991 JP 91P 139268, 12-JUL-1991 JP 91P 172794, PR 07-OCT-1991 JP 91P 287008, 16-DEC-1991 JP 91P 332329, PR 20-APR-1992 JP 92P 99957
PI SEKI MAKOTO, HONDA YOSHIKAZU, TAKAHASHI KAZUNOBU, PI MURAKAMI TOMOKO,
PI TERANISHI YUTAKA, HAYASHI NORIO
PC C12N15/51,C07K7/06,C07K7/10,C07K13/00,C07K15/12, PC C12N1/21,C12N5/10,
PC C12N15/11,C12N15/70,C12N15/85,C12P21/02//A61K39/00,A61K39/29, PC (C12R1:21,
PC (C12R1:19),(C12N5/10,C12R1:91),(C12P21/02,C12R1:19),(C12P21/02, PC C12R1:91),
PC C07K99:00;
CC strandedness: Double;
CC topology: Linear;
CC anti-sense: No;
CC *source: clone-MX25026A-1;
FH key Location/Qualifiers
FT mat_peptide 1..1278
FT /product='the peptides reacting specifically
FT and
FT immunochemically with the serum of hepatitis
FT

FT type C
FT patient',
FT Location/Qualifiers
FEATURES
source Location/Qualifiers
1..1280
/organism="Hepatitis C virus"
/db_xref="taxon:11103"
BASE COUNT 222 a 383 c 379 g 296 t
ORIGIN
Query Match 2.1%; Score 20; DB 6; Length 1280;
Best Local Similarity 100.0%; Pred. No. 24;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 311 CAGAGAACACAGAGAGA 330
|||||
Db 125 CAGAGAACACAGAGAGA 106
RESULT 34
E06217/c 1280 bp RNA linear PAT 29-SEP-1997
LOCUS
DEFINITION cDNA encoding genes derived from hepatitis C virus.
ACCESSION E06217
VERSION E06217.1 GI:2174404
KEYWORDS JP 1994000085-A/57.
SOURCE Hepatitis C virus.
ORGANISM Hepatitis C virus
Virus; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
Hepacivirus.
REFERENCE 1 (bases 1 to 1280)
AUTHORS Seki,M., Honda,Y., Takahashi,K., Murakami,T., Teranishi,Y. and Hayashi,N.
TITLE GENE OR DNA FRAGMENT DERIVED FROM HEPATITIS C VIRUS, POLYPEPTIDE CODED BY THE SAME AND ITS PRODUCTION
JOURNAL Patent: JP 1994000085-A 57 11-JAN-1994;
MITSUBISHI KASEI CORP
COMMENT OS (hepatitis C virus)
PN JP 1994000085-A/57
PD 11-JAN-1994
PE 11-JUN-1992 JP 1992194497
PR 11-JUN-1991 JP 91P 139268, 12-JUL-1991 JP 91P 172794, PR 07-OCT-1991 JP 91P 287008, 16-DEC-1991 JP 91P 332329, PR 20-APR-1992 JP 92P 99957
PI SEKI MAKOTO, HONDA YOSHIKAZU, TAKAHASHI KAZUNOBU, PI MURAKAMI TOMOKO,
PI TERANISHI YUTAKA, HAYASHI NORIO
PC C12N15/51,C07K7/06,C07K7/10,C07K13/00,C07K15/12, PC C12N1/21,C12N5/10,
PC C12N15/11,C12N15/70,C12N15/85,C12P21/02//A61K39/00,A61K39/29, PC (C12R1:21,
PC (C12R1:19),(C12N5/10,C12R1:91),(C12P21/02,C12R1:19),(C12P21/02, PC C12R1:91),
PC C07K99:00;
CC strandedness: Double;
CC topology: Linear;
CC anti-sense: No;
CC *source: clone-MX25026B-1;
FH key Location/Qualifiers
FT mat_peptide 1..1278
FT /product='the peptides reacting specifically
FT and
FT immunochemically with the serum of hepatitis
FT type C
FT patient',
FT Location/Qualifiers
FEATURES
source Location/Qualifiers
1..1280
/organism="Hepatitis C virus"
/db_xref="taxon:11103"
BASE COUNT 219 a 378 c 367 g 296 t
ORIGIN
Query Match 2.1%; Score 20; DB 6; Length 1280;

Best Local Similarity 100.0%; Pred. No. 24;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 311 CAGAGACACAGAGAAAGA 330
|||||
Db 125 CAGAGACACAGAGAAAGA 106

RESULT 35
E06396/C 1280 bp RNA linear PAT 29-SEP-1997
LOCUS E06396
DEFINITION CDNA encoding genes derived from hepatitis C virus.
ACCESSION E06396
VERSION E06396.1 GI:2174583
KEYWORDS JP 1994000086-A/40.
SOURCE Hepatitis C virus.
ORGANISM Hepatitis C virus.
REFERENCE 1 (bases 1 to 1280)
AUTHORS Seki,M., Honda,Y., Takahashi,K., Murakami,T., Teranishi,Y. and Hayashi,N.
TITLE GENE FOR DNA FRAGMENT DERIVED FROM HEPATITIS C VIRUS, POLYPEPTIDE CODED BY THE SAME AND ITS PRODUCTION
JOURNAL Patent: JP 1994000086-A 40 11-JAN-1994;
MITSUBISHI KASEI CORP
COMMENT OS (hepatitis C virus)
PN JP 1994000086-A/40
PD 11-JAN-1994
PF 07-OCT-1992 JP 1992293734
PR 07-OCT-1991 JP 91P 287008, 16-DEC-1991 JP 91P 332329, PR 20-APR-1992 JP 92P 99957
PI SEKI MAKOTO, HONDA YOSHIKAZU, TAKAHASHI KAZUNOBU, PI MURAKAMI TOMOKO,
PI TERANISHI YUTAKA, HAYASHI NORIO
PC C12N15/51,C07K7/06,C07K7/08,C07K7/10,C07K13/00,C12N5/10, PC C12N15/11,
PC C12N15/85,C12P21/02//A61K39/29,(C12P21/02,C12R1:91),C07K99:00;
CC strandedness: Double;
CC topology: Linear;
CC anti-sense: No;
CC *source: clone-MX25026;
FH Key Location/Qualifiers
FT mat_peptide 1..1278
FT /product='the peptides reacting specifically and immunochemically with the serum of hepatitis type C patient'.
FT Location/Qualifiers
FT 1..1280
FEATURES
source 1..1280
BASE COUNT 203 a 354 c 356 g 269 t 98 others
ORIGIN
Query Match 2.1%; Score 20; DB 6; Length 1280;
Best Local Similarity 100.0%; Pred. No. 24;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 311 CAGAGACACAGAGAAAGA 330
|||||
Db 125 CAGAGACACAGAGAAAGA 106
RESULT 36
E06412 1280 bp RNA linear PAT 29-SEP-1997
LOCUS E06412
DEFINITION CDNA encoding genes derived from hepatitis C virus.
ACCESSION E06412.1 GI:2174599
VERSION E06412.1
KEYWORDS JP 1994000086-A/56.

SOURCE Hepatitis C virus.
ORGANISM Hepatitis C virus
VIRUSES: ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
Hepacivirus.

REFERENCE 1 (bases 1 to 1280)
AUTHORS Seki,M., Honda,Y., Takahashi,K., Murakami,T., Teranishi,Y. and Hayashi,N.
TITLE GENE FOR DNA FRAGMENT DERIVED FROM HEPATITIS C VIRUS, POLYPEPTIDE CODED BY THE SAME AND ITS PRODUCTION
JOURNAL Patent: JP 1994000086-A 56 11-JAN-1994;
MITSUBISHI KASEI CORP
COMMENT OS (hepatitis C virus)
PN JP 1994000086-A/56
PD 11-JAN-1994
PF 07-OCT-1992 JP 1992293734
PR 07-OCT-1991 JP 91P 287008, 16-DEC-1991 JP 91P 332329, PR 20-APR-1992 JP 92P 99957
PI SEKI MAKOTO, HONDA YOSHIKAZU, TAKAHASHI KAZUNOBU, PI MURAKAMI TOMOKO,
PI TERANISHI YUTAKA, HAYASHI NORIO.
PC C12N15/51,C07K7/06,C07K7/08,C07K7/10,C07K13/00,C12N5/10, PC C12N15/11,
PC C12N15/85,C12P21/02//A61K39/29,(C12P21/02,C12R1:91),C07K99:00;
CC strandedness: Double;
CC topology: Linear;
CC anti-sense: No;
CC *source: clone-MX25026A-1;
FH Key Location/Qualifiers
FT mat_peptide 1..1278
FT /product='the peptides reacting specifically and immunochemically with the serum of hepatitis type C patient'.
FT Location/Qualifiers
FT 1..1280
FEATURES
source 1..1280
BASE COUNT 222 a 383 c 379 g 296 t
ORIGIN
Query Match 2.1%; Score 20; DB 6; Length 1280;
Best Local Similarity 100.0%; Pred. No. 24;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 311 CAGAGACACAGAGAAAGA 330
|||||
Db 125 CAGAGACACAGAGAAAGA 106
RESULT 37
E06413 1280 bp RNA linear PAT 29-SEP-1997
LOCUS E06413/C
DEFINITION CDNA encoding genes derived from hepatitis C virus.
ACCESSION E06413
VERSION E06413.1 GI:2174600
KEYWORDS JP 1994000086-A/57.
SOURCE Hepatitis C virus.
ORGANISM Hepatitis C virus.
VIRUSES: ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
Hepacivirus.

REFERENCE 1 (bases 1 to 1280)
AUTHORS Seki,M., Honda,Y., Takahashi,K., Murakami,T., Teranishi,Y. and Hayashi,N.
TITLE GENE FOR DNA FRAGMENT DERIVED FROM HEPATITIS C VIRUS, POLYPEPTIDE CODED BY THE SAME AND ITS PRODUCTION
JOURNAL Patent: JP 1994000086-A 57 11-JAN-1994;
MITSUBISHI KASEI CORP
COMMENT OS (hepatitis C virus)
PN JP 1994000086-A/57
PD 11-JAN-1994
PF 07-OCT-1992 JP 1992293734

PR 07-OCT-1991 JP 91P 287008, 16-DEC-1991 JP 91P 332329, PR
 20-APR-1992 JP 92P 99957
 PI SEKI MAKOTO, HONDA YOSHIKAZU, TAKAHASHI KAZUNOBU, PI
 MURAKAMI TOMOKO,
 PI TERANISHI YUTAKA, HAYASHI NORIO
 PC C12N15/51, C07K7/06, C07K7/10, C07K13/00, C12N5/10, PC
 C12N15/11,
 PC C12N15/85, C12P21/02//A61K39/29, (C12P21/02, C12R1:91), C07K99:00;
 CC strandedness: Double;
 CC topology: Linear;
 CC anti-sense: No;
 CC *source: clone-MX25026B-1;
 FH key Location/Qualifiers
 FT mat_peptide 1..1278
 FT /product='the peptides reacting specifically
 FT and
 FT immunohchemically with the serum of hepatitis
 FT type C
 FT patient',
 FT
 FT
 FT
 FT

FEATURES
 source Location/Qualifiers
 1..1280
 /organism="Hepatitis C virus"
 /db_xref="taxon:11103"

BASE COUNT 219 a 378 c 387 g 296 t

ORIGIN

Query Match 2.1%; Score 20; DB 6; Length 1280;
 Best Local Similarity 100.0%; Pred. No. 24;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 311 CAGAGAACACAGGAAGA 330
 ||||||||||||||||||
 DB 125 CAGAGAACACAGGAAGA 106

RESULT 38
 E06203/c 3564 bp RNA linear PAT 29-SEP-1997
 LOCUS E06203
 DEFINITION cDNA encoding genes derived from hepatitis C virus.
 ACCESSION E06203
 VERSION E06203.1 GI:2174390
 KEYWORDS JP 1994000085-A/43.
 SOURCE Hepatitis C virus.
 ORGANISM Hepatitis C virus
 viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 Hepacivirus.
 1 (bases 1 to 3564)
 Seki,M., Honda,Y., Takahashi,K., Murakami,T., Teranishi,Y. and
 Hayashi,N.
 GENE OR DNA FRAGMENT DERIVED FROM HEPATITIS C VIRUS, POLYPEPTIDE
 CODED BY THE SAME AND ITS PRODUCTION
 Patent: JP 1994000085-A 43 11-JAN-1994;
 JOURNAL MITSUBISHI KASEI CORP
 OS (hepatitis C virus)
 PN JP 1994000085-A/43
 PD 11-JUN-1994
 PF 11-JUN-1992 JP 1992194497
 PR 11-JUN-1991 JP 91P 139268, 12-JUL-1991 JP 91P 172794, PR
 07-OCT-1991 JP 91P 287008, 16-DEC-1991 JP 91P 332329, PR
 20-APR-1992 JP 92P 99957
 PI SEKI MAKOTO, HONDA YOSHIKAZU, TAKAHASHI KAZUNOBU, PI
 MURAKAMI TOMOKO,
 PI TERANISHI YUTAKA, HAYASHI NORIO
 PC C12N15/51, C07K7/06, C07K7/10, C07K13/00, C07K15/12, PC
 C12N1/21, C12N5/10,
 PC C12N15/11, C12N15/70, C12N15/85, C12P21/02//A61K39/00, A61K39/29,
 PC (C12N1/21,
 PC C12R1:19), (C12N5/10, C12R1:91), (C12P21/02, C12R1:19), (C12P21/02,
 PC C07K99:00;
 CC strandedness: Double;
 CC topology: Linear;

CC anti-sense: No;
 CC *source: clone-MX25N15;
 FH key Location/Qualifiers
 FH mat_peptide 1..3564
 FH /product='the peptides reacting specifically
 FH and
 FH immunohchemically with the serum of hepatitis
 FH type C
 FH patient',
 FH
 FH
 FH
 FH

FEATURES
 source Location/Qualifiers
 1..3564
 /organism="Hepatitis C virus"
 /db_xref="taxon:11103"

BASE COUNT 640 a 1019 c 1024 g 747 t 124 others

ORIGIN

QY 311 CAGAGAACACAGGAAGA 330
 ||||||||||||||||||
 DB 125 CAGAGAACACAGGAAGA 106

RESULT 39
 E06223/c 3564 bp RNA linear PAT 29-SEP-1997
 LOCUS E06223
 DEFINITION cDNA encoding genes derived from hepatitis C virus.
 ACCESSION E06223
 VERSION E06223.1 GI:2174410
 KEYWORDS JP 1994000085-A/63.
 SOURCE Hepatitis C virus.
 ORGANISM Hepatitis C virus
 viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 Hepacivirus.
 1 (bases 1 to 3564)
 Seki,M., Honda,Y., Takahashi,K., Murakami,T., Teranishi,Y. and
 Hayashi,N.
 GENE OR DNA FRAGMENT DERIVED FROM HEPATITIS C VIRUS, POLYPEPTIDE
 CODED BY THE SAME AND ITS PRODUCTION
 Patent: JP 1994000085-A 63 11-JAN-1994;
 JOURNAL MITSUBISHI KASEI CORP
 OS (hepatitis C virus)
 PN JP 1994000085-A/63
 PD 11-JAN-1994
 PF 11-JUN-1992 JP 1992194497
 PR 11-JUN-1991 JP 91P 139268, 12-JUL-1991 JP 91P 172794, PR
 07-OCT-1991 JP 91P 287008, 16-DEC-1991 JP 91P 332329, PR
 20-APR-1992 JP 92P 99957
 PI SEKI MAKOTO, HONDA YOSHIKAZU, TAKAHASHI KAZUNOBU, PI
 MURAKAMI TOMOKO,
 PI TERANISHI YUTAKA, HAYASHI NORIO
 PC C12N15/51, C07K7/06, C07K7/10, C07K13/00, C07K15/12, PC
 C12N1/21, C12N5/10,
 PC C12N15/11, C12N15/70, C12N15/85, C12P21/02//A61K39/00, A61K39/29,
 PC (C12N1/21,
 PC C12R1:19), (C12N5/10, C12R1:91), (C12P21/02, C12R1:19), (C12P21/02,
 PC C07K99:00;
 CC strandedness: Double;
 CC topology: Linear;
 CC anti-sense: No;
 CC *source: clone-MX25N15-1;
 FH key Location/Qualifiers
 FH mat_peptide 1..3564
 FH /product='the peptides reacting specifically
 FH and
 FH immunohchemically with the serum of hepatitis
 FH type C
 FH patient',
 FH
 FH
 FH
 FH

FEATURES
source 1. .3564
Location/Qualifiers
/organism="Hepatitis C virus"
/db_xref="taxon:11103"

BASE COUNT 664 a 1062 c 1055 g 783 t

ORIGIN

Query Match 2.1%; Score 20; DB 6; Length 3564;
Best Local Similarity 100.0%; Pred. No. 23;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 311 CAGAGAACACAGGAAGA 330
|||||

Db 125 CAGAGAACACAGGAAGA 106

RESULT 40
E06399/c 3564 bp RNA linear PAT 29-SEP-1997

LOCUS E06399
DEFINITION cDNA encoding genes derived from hepatitis C virus.
ACCESSION E06399.1 GI:2174586
VERSION JP 1994000086-A/43.
KEYWORDS Hepatitis C virus.
SOURCE Hepatitis C virus.
ORGANISM Hepatitis C virus.
Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae; Hepacivirus.

REFERENCE 1 (bases 1 to 3564)
Seki,M., Honda,Y., Takahashi,K., Murakami,T., Teranishi,Y. and Hayashi,N.
GENE FOR DNA FRAGMENT DERIVED FROM HEPATITIS C VIRUS, POLYPEPTIDE CODED BY THE SAME AND ITS PRODUCTION
Patent: JP 1994000086-A 43 11-JAN-1994;
MITSUBISHI KASEI CORP
OS (hepatitis C virus)
PN JP 1994000086-A/43
PD 11-JAN-1994
PF 07-OCT-1992 JP 1992293734
PR 07-OCT-1991 JP 91P 287008, 16-DEC-1991 JP 91P 332329, PR 20-APR-1992 JP 92P 99957
PI SEKI MAKOTO, HONDA YOSHIKAZU, TAKAHASHI KAZUNOBU, PI MURAKAMI TOMOKO,
PI TERANISHI YUTAKA, HAYASHI NORIO
PC C12N15/51,C07K7/06,C07K7/10,C07K13/00,C12N5/10, PC C12N15/11,
PC C12N15/85,C12P21/02//A61K39/29,(C12P21/02,C12R1:91),C07K99:00;
CC strandedness: Double;
CC topology: Linear;
CC anti-sense: No;
CC *source: clone-MX25N15;
FH Key Location/Qualifiers
FT mat_peptide 1. .3564
FT /product="the peptides reacting specifically
FT and
FT immunochemically with the serum of hepatitis
FT type C
FT patient".
FT Location/Qualifiers

FEATURES
source 1. .3564
Location/Qualifiers
/organism="Hepatitis C virus"
/db_xref="taxon:11103"

BASE COUNT 640 a 1019 c 1024 g 747 t 134 others

ORIGIN

Query Match 2.1%; Score 20; DB 6; Length 3564;
Best Local Similarity 100.0%; Pred. No. 23;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 311 CAGAGAACACAGGAAGA 330
|||||

Db 125 CAGAGAACACAGGAAGA 106

RESULT 41
E06419/c 3564 bp RNA linear PAT 29-SEP-1997

LOCUS E06419
DEFINITION cDNA encoding genes derived from hepatitis C virus.
ACCESSION E06419.1 GI:2174606
VERSION JP 1994000086-A/63.
KEYWORDS Hepatitis C virus.
SOURCE Hepatitis C virus.
ORGANISM Hepatitis C virus.
Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae; Hepacivirus.

REFERENCE 1 (bases 1 to 3564)
Seki,M., Honda,Y., Takahashi,K., Murakami,T., Teranishi,Y. and Hayashi,N.
GENE FOR DNA FRAGMENT DERIVED FROM HEPATITIS C VIRUS, POLYPEPTIDE CODED BY THE SAME AND ITS PRODUCTION
Patent: JP 1994000086-A 63 11-JAN-1994;
MITSUBISHI KASEI CORP
OS (hepatitis C virus)
PN JP 1994000086-A/63
PD 11-JAN-1994
PF 07-OCT-1992 JP 1992293734
PR 07-OCT-1991 JP 91P 287008, 16-DEC-1991 JP 91P 332329, PR 20-APR-1992 JP 92P 99957
PI SEKI MAKOTO, HONDA YOSHIKAZU, TAKAHASHI KAZUNOBU, PI MURAKAMI TOMOKO,
PI TERANISHI YUTAKA, HAYASHI NORIO
PC C12N15/51,C07K7/06,C07K7/10,C07K13/00,C12N5/10, PC C12N15/11,
PC C12N15/85,C12P21/02//A61K39/29,(C12P21/02,C12R1:91),C07K99:00;
CC strandedness: Double;
CC topology: Linear;
CC anti-sense: No;
CC *source: clone-MX25N15-1;
FH Key Location/Qualifiers
FT mat_peptide 1. .3564
FT /product="the peptides reacting specifically
FT and
FT immunochemically with the serum of hepatitis
FT type C
FT patient".
FT Location/Qualifiers

FEATURES
source 1. .3564
Location/Qualifiers
/organism="Hepatitis C virus"
/db_xref="taxon:11103"

BASE COUNT 664 a 1062 c 1055 g 783 t

ORIGIN

Query Match 2.1%; Score 20; DB 6; Length 3564;
Best Local Similarity 100.0%; Pred. No. 23;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 311 CAGAGAACACAGGAAGA 330
|||||

Db 125 CAGAGAACACAGGAAGA 106

RESULT 42
AF313916/c 9359 bp mRNA linear VRL 01-JAN-2002

LOCUS AF313916
DEFINITION Hepatitis C virus polyprotein mRNA, partial cds.
ACCESSION AF313916
VERSION AF313916.1 GI:18027684
KEYWORDS
SOURCE Hepatitis C virus.
ORGANISM Hepatitis C virus.
Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae; Hepacivirus.

REFERENCE 1 (bases 1 to 9359)
Fanning,L.J., Itakura,J., Nagayama,K. and Enomoto,N.
Characteristics of Hepatitis C viral genome associated with disease

JOURNAL Unpublished
REFERENCE 2 (bases 1 to 9359)
AUTHORS Fanning, L. J., Itakura, J., Nagayama, K. and Enomoto, N
TITLE Direct Submission
JOURNAL Submitted (17-OCT-2000) Medicine, National University of
Clinical Sciences Building

PVPSDMETKITTGWADTAACGDIILGLPVSARCKEILLGPADSLGQWMLLAPIT
 AYSOQTRGLGCIITSLGRKNOVEGEVOVSTQSLACINGVCTVHGAGSK
 TLAGRGCTITQWTVNDDVLGMPAPGASMTPTCTGSLTITRADVIVPRRG
 DSRGSLSPRISTYLSKSGGGLPCSGHVVAIFRAACTRGVAAVFPVPSMETT
 MRSPTFTDNSTPPAPYQSSQVHLLHAPGSSKSTVPAPYAAOGKIVLVNLSVATL
 GGAAYMSKAHVDPNIRTVGRTITTGAPITVSTYKFLADGSGGAGDIILICDCHS
 TDSTSLIGITVLDOAETAGARLVLAATAPGSAVPPNIEVALNPTSGDIIPIYG
 AIPLEAIGKRLHIFCHSKKCDLAALASAGVANAVAYRLDVSIIPTSGDVVVA
 TDLMTGYTGDDESDVINDCTVOTGVDSLDPTFTETTVTVDRVRSORGRPRG
 RGCIATRYTPGRSGMFDSSYLCECYDAGCMVLTAEVTSRIATYINTGDLVCO
 DHEFWESFTGLTHIDAHFSLQTKOAGDNFYLVAIQATVCARAKAPPSSDWMKC
 LIRLKPPLHGPPLLYRLGAVONEVTLHPITKFLMACMSADLEVSTWVLVGVLA
 ALAACTLTGTVIVIGRIILSGRPVVIDREVLXOEPMEECASHLPYIEGMLAE
 OFKOKALILQATKOAEAAAVVSESKRALETFWAKHMMNFISGTOYLAGSTLPGN
 PAISALMFTASTISPLTTOYTLLENIIGWVAOLAPPSASAFVAGIACAAGSI
 GIGKYLINDIAGYAGVAGALVAFKVMGDMSTEDVNLPAIISPCALYGVVCA
 ILRRVYGEBAVOMNMLIFASRGHNSPHIVPESDARAVTQIISNLITDOLK
 RLHOWINDECSPTCSGSLRDVMDIICVVLADFKTWLQSLPLRGVPEFSCQRYK
 GVRGDIIMYITPCGAQITGHVKNKSMRIVGPRCSMTMGTPTINATYTPGCTPSP
 APNYSRALMRVAAEEVEVTRVGDHNYVTGMTDNKPCQVPAPEFTELDGVILHR
 YAPACKPLRDEVTQVGLNVTGVSOLCEPEPDVVTSMITDPSHTTAARRL
 ARGSPSLASSASAOLSLKATCTTHGAPCDPLIFANILMROEMGNTIRVSEN
 KIVLIDSEPLAEEDEREVVAETILKTRKFPAMVWAPRDINPPLLESKAPDY
 VPPVHGGPLPEPTKAPPIPPRRKRTVVLTESTVSSALAEATKTFGSSGSAVDST
 ATGPDQASAECDAGSDAESYSSMPLEGEPPDPLNDGSMSTVEEASSEDVCCMS
 YTMGTALITPCAAEESKLPIINALSNPLRHHNMVATTSRSASQOKKVTPELRLVD
 DHYRDVLDMKAKASTVAKLLSYEACKLTPPHSARSKFGAGKDVRSLSKAVNI
 RSVMKLIDEDFTPTIDTITMAKNEVCEQPEKGRPARLIVFPDLGVAYCEKMAVLD
 VYSTLPQAVMSSYGFQYSPGQRFVFLVNAKSKSPMGFAIDTCCPSTYTENDIRT
 EESTIYOCDDLEPAAQAIRSLTERLYIGGPLNKSQONCGIRKCVSGLVTTSCNTL
 TCYLKASACRAKALQDCTMLVCGDVLVYICDSAGTODASLRVETAMTRYSAAPG
 DPPOEYDLELITSCSSNVSAHDASGRVYLLDPTPTPLARAMEYARHPVNSML
 GNIIWAPTLMARMLTMHFFSIIILAOLEKALDQOIGATYISIEPLDLPDIORLH
 GLSAFSLSYSGEINRYVASCRLKGLVPLRYMRARSVRAKILISOGGRATCKYL
 FNMAYKTKLKLPIPEASQDLSGMFVAGYSGGDIYHLSLRARPPWPMCLLISVY
 GYILPNR"

misc_feature
 3029..4490
 /note="Csp1-XhoI(MKc1a clone), other sequence(MKc1c clone)"

BASE COUNT 1893 a 2820 c 2693 g 2025 t
 ORIGIN

Query Match 2.1%; Score 20; DB 14; Length 9431;
 Best Local Similarity 100.0%; Pred. No. 22;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 311 CAGAGAACACAAAGAAAGA 330
 Db 2659 CAGAGAACACAAAGAAAGA 2640

RESULT 44
 LOCUS HPCPP 9431 bp RNA linear VRL 07-FEB-1999
 DEFINITION Hepatitis C virus complete genome sequence.
 ACCESSION D30613.1 GI:560788
 VERSION D30613.1 GI:560788
 KEYWORDS polyprotein.
 SOURCE Hepatitis C virus cDNA to genomic RNA.
 ORGANISM Hepatitis C virus
 Viruses: ssRNA positive-strand viruses, no DNA stage; Flaviviridae; Hepacivirus.

REFERENCE 1 (sites)
 AUTHORS Seki,M., Honda,Y., Kondo,J., Fukuda,K., Ohta,K., Sugimoto,J. and Yamada,E.
 TITLE Effective production of the hepatitis C virus core antigen having high purity in *Escherichia coli*
 JOURNAL J. Biotechnol. 38 (3), 229-241 (1995)
 MEDLINE 95186115
 REFERENCE 2 (bases 1 to 9431)
 AUTHORS Seki,M. and Honda,Y.
 JOURNAL Unpublished
 REFERENCE 3 (bases 1 to 9431)

AUTHORS Seki,M.
 TITLE Direct Submission
 JOURNAL Submitted (29-Apr-1994) Makoto Seki, Mitsubishi Chemical Corporation Yokohama Research Center, Department of Biotechnology, 1000 Kamoshida-cho, Koda-ku, Yokohama, Kanagawa 227, Japan (E-mail:seki@atlas.ic.m.kagaku.co.jp, Tel:045-963-3455, Fax:045-963-3992)
 COMMENT The serums of a chronic HCV patients were gifted by Dr. N. Hayashi and Dr. E. Mita(Osaka University Medical School).

FEATURES
 source Location/Qualifiers

5'UTR
 CDS

1..9431
 /organism="Hepatitis C virus"
 /db_xref="taxon:11103"
 1..341
 342..9374
 /codon_start=1
 /product="polyprotein"
 /protein_id="BAA06303.1"
 /db_xref="GI:560789"
 /translation="MSTNPKPQRIKRNTRRRPQVKKPPGGQIYGVYLLPRGRPLGVNATGNLPQCSFSIFLLALISCLTIPASAVQVNRASGVHYVTDSCSSSTVYADVYMHDPGCVNRENNSRCVWALPTLAARNSLPTTIRRHVDLLVGAALCSAMYVDFGSGVFLVSLQTFEPRRYETQDCSTYPGVSGHRAMAMIMNMSPVTAUVSOLLRIPOAVDMVAGAHMGVLAGLAAYSMVCMMAKVLVMLLEAGVDGCTVTGKVAATGQFTSEFSRSPGSKTOIVNTGSMHINRTALNCNLSLNTGFLAALFYHSPNASGCPREMACRGITDEFACQSGPLPCPSRVNGIFRAACTRGVAAVFPVPSMETTFTPSVYVGTDRFGPTTNMGNNETDVLNNTNRPQGMNGCTMMNGTFTKTCGGPCNIGVGNNTLCTPDCFRHPEATYVTKCSGSMPLRCLVHPYRLMHPKCVNFTEFKVIRMGVGEHRLHACNMTREGRDLEDRSELSPLISTEMOVLPCSEFTLPALSTGLIHGHONIYDVQYLYIGSAVFAIKWEYILLPLFLDARACAMLMMLIAOAAALENVLVNAAAGAGVGLVFLFSYKFKLIPKLVLMWLOYLITVRAELLIMALPARVANDREMAASGCAVGVGLVFLFSYKFKLIPKLVLMWLOYLITVRAELLQWVPLPLNRCGRALITLLCAVHPELFDITKLLALILGLPLMVLQGLTQMPYFVRAOGLIRMCCLVRAAGHYVQMAKMLAALTSTYVYDHLTPLDMAHARLDLAVAEPVFSQTRGLGCIITSLGRKNOVEGEVOVSTQSLACINGVCTVHGAGSKTLAGRGCTITQWTVNDDVLGMSAPGASRLTPCQSGSDLYLVTRADVIVPRRGMSRPTFTDNSTPPAPYQSSQVHLLHAPGSSKSTVPAPYAAOGKIVLVNLSVATLGGAYMSKAHVDPNIRTVGRTITTGAPITVSTYKFLADGSGGAGDIILICDCHSADSTSLIGITVLDOAETAGARLVLAATAPGSAVPPNIEVALNPTSGDIIPIYGAIPLAIGKRLHIFCHSKKCDLAALASAGVANAVAYRLDVSIIPTSGDVVVA TDLMTGYTGDDESDVINDCTVOTGVDSLDPTFTETTVTVDRVRSORGRPRG RGCIATRYTPGRSGMFDSSYLCECYDAGCMVLTAEVTSRIATYINTGDLVCO DHEFWESFTGLTHIDAHFSLQTKOAGDNFYLVAIQATVCARAKAPPSSDWMKC LIRLKPPLHGPPLLYRLGAVONEVTLHPITKFLMACMSADLEVSTWVLVGVLA ALAACTLTGTVIVIGRIILSGRPVVIDREVLXOEPMEECASHLPYIEGMLAE OFKOKALILQATKOAEAAAVVSESKRALETFWAKHMMNFISGTOYLAGSTLPGN PAISALMFTASTISPLTTOYTLLENIIGWVAOLAPPSASAFVAGIACAAGSI GIGKYLINDIAGYAGVAGALVAFKVMGDMSTEDVNLPAIISPCALYGVVCA ILRRVYGEBAVOMNMLIFASRGHNSPHIVPESDARAVTQIISNLITDOLK RLHOWINDECSPTCSGSLRDVMDIICVVLADFKTWLQSLPLRGVPEFSCQRYK GVRGDIIMYITPCGAQITGHVKNKSMRIVGPRCSMTMGTPTINATYTPGCTPSP APNYSRALMRVAAEEVEVTRVGDHNYVTGMTDNKPCQVPAPEFTELDGVILHR YAPACKPLRDEVTQVGLNVTGVSOLCEPEPDVVTSMITDPSHTTAARRL ARGSPSLASSASAOLSLKATCTTHGAPCDPLIFANILMROEMGNTIRVSEN KIVLIDSEPLAEEDEREVVAETILKTRKFPAMVWAPRDINPPLLESKAPDY VPPVHGGPLPEPTKAPPIPPRRKRTVVLTESTVSSALAEATKTFGSSGSAVDST ATGPDQASAECDAGSDAESYSSMPLEGEPPDPLNDGSMSTVEEASSEDVCCMS YTMGTALITPCAAEESKLPIINALSNPLRHHNMVATTSRSASQOKKVTPELRLVD DHYRDVLDMKAKASTVAKLLSYEACKLTPPHSARSKFGAGKDVRSLSKAVNI RSVMKLIDEDFTPTIDTITMAKNEVCEQPEKGRPARLIVFPDLGVAYCEKMAVLD VYSTLPQAVMSSYGFQYSPGQRFVFLVNAKSKSPMGFAIDTCCPSTYTENDIRT EESTIYOCDDLEPAAQAIRSLTERLYIGGPLNKSQONCGIRKCVSGLVTTSCNTL TCYLKASACRAKALQDCTMLVCGDVLVYICDSAGTODASLRVETAMTRYSAAPG DPPOEYDLELITSCSSNVSAHDASGRVYLLDPTPTPLARAMEYARHPVNSML GNIIWAPTLMARMLTMHFFSIIILAOLEKALDQOIGATYISIEPLDLPDIORLH GLSAFSLSYSGEINRYVASCRLKGLVPLRYMRARSVRAKILISOGGRATCKYL FNMAYKTKLKLPIPEASQDLSGMFVAGYSGGDIYHLSLRARPPWPMCLLISVY GYILPNR"

BASE COUNT 1882 a 2819 c 2701 g 2029 t

ORIGIN

Query Match 2.1%; Score 20; DB 14; Length 9431;
Best Local Similarity 100.0%; Pred. No. 22;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 311 CAGAGAACACAGGAAGA 330
|||||
DB 2659 CAGAGAACACAGGAAGA 2640

RESULT 45

E06261/c

LOCUS E06261 9471 bp RNA linear PAT 29-SEP-1997
DEFINITION cDNA encoding genes derived from hepatitis C virus.
ACCESSION E06261.1 GI:2174448
VERSION JP 1994000085-A/101.
KEYWORDS Hepatitis C virus.
SOURCE Hepatitis C virus.
ORGANISM Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;

REFERENCE

1 (bases 1 to 9471)
Seki,M., Honda,Y., Takahashi,K., Murakami,T., Teranishi,Y. and
Hayashi,N.
GENE OR DNA FRAGMENT DERIVED FROM HEPATITIS C VIRUS, POLYPEPTIDE
CODED BY THE SAME AND ITS PRODUCTION
Patent: JP 1994000085-A 101 11-JAN-1994;
MITSUBISHI KASEI CORP
OS (hepatitis C virus)
PN JP 1994000085-A/101
PD 11-JAN-1994
PF 11-JUN-1992 JP 1992194497
PR 11-JUN-1991 JP 91P 139268, 12-JUL-1991 JP 91P 172794, PR
07-OCT-1991 JP 91P 287008, 16-DEC-1991 JP 91P 332329, PR
20-APR-1992 JP 92P 99957
PI SEKI MAKOTO, HONDA YOSHIKAZU, TAKAHASHI KAZUNOBU, PI
MURAKAMI TOMOKO,
PI TERANISHI YUTAKA, HAYASHI NORIO
PC C12N15/51,C07K7/06,C07K7/10,C07K13/00,C07K15/12, PC
C12N1/21,C12N5/10,
PC C12N15/11,C12N15/70,C12N15/85,C12P21/02//A61K39/00,A61K39/29,
PC (C12N1/21,
PC C12R1:19),(C12N5/10,C12R1:91),(C12P21/02,C12R1:19),(C12P21/02,
PC C12R1:91),
PC C07K99:00;
CC strandedness: Double;
CC topology: Linear;
CC anti-sense: No;
CC *source: clone=T7N1-30;
FH Key Location/Qualifiers
FH mat-peptide 369..9398
FT /product='the peptides reacting specifically
FT and
FT immunochemically with the serum of hepatitis
FT type C
FT patient',
FT

COMMENT

JOURNAL

TITLE

GENE OR DNA FRAGMENT DERIVED FROM HEPATITIS C VIRUS, POLYPEPTIDE
CODED BY THE SAME AND ITS PRODUCTION
Patent: JP 1994000085-A 101 11-JAN-1994;
MITSUBISHI KASEI CORP
OS (hepatitis C virus)
PN JP 1994000085-A/101
PD 11-JAN-1994
PF 11-JUN-1992 JP 1992194497
PR 11-JUN-1991 JP 91P 139268, 12-JUL-1991 JP 91P 172794, PR
07-OCT-1991 JP 91P 287008, 16-DEC-1991 JP 91P 332329, PR
20-APR-1992 JP 92P 99957
PI SEKI MAKOTO, HONDA YOSHIKAZU, TAKAHASHI KAZUNOBU, PI
MURAKAMI TOMOKO,
PI TERANISHI YUTAKA, HAYASHI NORIO
PC C12N15/51,C07K7/06,C07K7/10,C07K13/00,C07K15/12, PC
C12N1/21,C12N5/10,
PC C12N15/11,C12N15/70,C12N15/85,C12P21/02//A61K39/00,A61K39/29,
PC (C12N1/21,
PC C12R1:19),(C12N5/10,C12R1:91),(C12P21/02,C12R1:19),(C12P21/02,
PC C12R1:91),
PC C07K99:00;
CC strandedness: Double;
CC topology: Linear;
CC anti-sense: No;
CC *source: clone=T7N1-30;
FH Key Location/Qualifiers
FH mat-peptide 369..9398
FT /product='the peptides reacting specifically
FT and
FT immunochemically with the serum of hepatitis
FT type C
FT patient',
FT

FEATURES

Source

BASE COUNT 1892 a 2830 c 2712 g 2037 t
ORIGIN

Query Match 2.1%; Score 20; DB 6; Length 9471;
Best Local Similarity 100.0%; Pred. No. 22;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 311 CAGAGAACACAGGAAGA 330
|||||
DB 2686 CAGAGAACACAGGAAGA 2667

RESULT 46

E06457/c

LOCUS E06457 9471 bp RNA linear PAT 29-SEP-1997
DEFINITION cDNA encoding genes derived from hepatitis C virus.
ACCESSION E06457
VERSION E06457.1 GI:2174644
KEYWORDS JP 1994000086-A/101.
SOURCE Hepatitis C virus.
ORGANISM Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;

REFERENCE

1 (bases 1 to 9471)
Seki,M., Honda,Y., Takahashi,K., Murakami,T., Teranishi,Y. and
Hayashi,N.
GENE FOR DNA FRAGMENT DERIVED FROM HEPATITIS C VIRUS, POLYPEPTIDE
CODED BY THE SAME AND ITS PRODUCTION
Patent: JP 1994000086-A 101 11-JAN-1994;
MITSUBISHI KASEI CORP
OS (hepatitis C virus)
PN JP 1994000086-A/101
PD 11-JAN-1994
PF 07-OCT-1992 JP 1992293734
PR 07-OCT-1991 JP 91P 287008, 16-DEC-1991 JP 91P 332329, PR
20-APR-1992 JP 92P 99957
PI SEKI MAKOTO, HONDA YOSHIKAZU, TAKAHASHI KAZUNOBU, PI
MURAKAMI TOMOKO,
PI TERANISHI YUTAKA, HAYASHI NORIO
PC C12N15/51,C07K7/06,C07K7/10,C07K13/00,C12N5/10, PC
C12N15/11,
PC C12N15/85,C12P21/02//A61K39/29,(C12P21/02,C12R1:91),C07K99:00;
CC strandedness: Double;
CC topology: Linear;
CC anti-sense: No;
CC *source: clone=T7N1-30;
FH Key Location/Qualifiers
FH mat-peptide 369..9398
FT /product='the peptides reacting specifically
FT and
FT immunochemically with the serum of hepatitis
FT type C
FT patient',
FT

JOURNAL

COMMENT

BASE COUNT 1892 a 2830 c 2712 g 2037 t
ORIGIN

Query Match 2.1%; Score 20; DB 6; Length 9471;
Best Local Similarity 100.0%; Pred. No. 22;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 311 CAGAGAACACAGGAAGA 330
|||||
DB 2686 CAGAGAACACAGGAAGA 2667

RESULT 47

AB049099/c

LOCUS AB049099 9504 bp RNA linear VRL 22-AUG-2002
DEFINITION Hepatitis C virus gene for polyprotein, complete cds,
Isolate:HCVT212.
AB049099
ACCESSION AB049099.1 GI:11559464
VERSION
KEYWORDS Hepatitis C virus (isolate:HCVT212, specific_host:human) serum cDNA
to genomic RNA.
SOURCE Hepatitis C virus
ORGANISM Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
REFERENCE 1

AUTHORS Takahashi, K., Iwata, K., Matsumoto, M., Matsumoto, H., Nakao, K., Hatahara, T., Ohta, Y., Kanai, K., Maruo, H., Baba, K., Hijikata, M. and Mishiro, S.

TITLE Hepatitis C virus (HCV) genotype 1b sequences from fifteen patients with hepatocellular carcinoma: the 'progression score' revisited

JOURNAL Hepatol. Res. 20 (2), 161-171 (2001)

PMID 11348851

REFERENCE 2 (bases 1 to 9504)

AUTHORS Mishiro, S.

JOURNAL Direct Submission

Submitted (19-SEP-2000) Shunji Mishiro, Toshiba General Hospital, Department of Medical Sciences, 6-3-22 Higashi Oh-1, Shingawa-ku, Tokyo 140-8522, Japan (E-mail: shunji.mishiro@po.toshiba.co.jp, Tel:81-3-3764-8981, Fax:81-3-3764-8992)

FEATURES

source

1. 9504

/organism="Hepatitis C virus"

/isolate="HCV1212"

/specific_host="human"

/db_xref="taxon:11103"

/tissue_type="serum"

/note="HCV-genotype: 1b"

1. 341

/note="complete"

342. 9377

/codon_start=1

/product="polyprotein"

/protein_id="BAB18812.1"

/db_xref="GI:11559465"

/translation="MSTNPKPQRTKRNTRRPODVKPPGGQIVGVYLLPRGPRRLGPRVTRKTSERQPRGRROPIPKARRPGRRAAGYGMPLVGNMGMAWGLAPRGSPRWGPTDPRRRRNKGVLDITLGGADLADYIPYGAAPGAAARALAHGVLEADGVNATGMLPGCSFSLFLLALSLCLIPASAEVNVSGGYVNTDNCSSSYVETADMIMHPTGVCVCRSNTSRWCWALPPLTARNSNPTMIRHIDILVGTARFCAMVCDICGSEYVNSQLTFPSRREHYVDCNCSYSPHLHGHRAAMMMAMMSTALVYV SOLLRIPQAVDMVAGAHGVLAGLAVYSWGNNAKVLIVMLFAGVGRIGLIGGTSINTEVVASIESPSPQKIQLVNINGSWIMINTALDCKASLQTCGLAGLIVAKRNTSGLEPMASCRPIDTFAOGMGPITTYERVNSDQPYCMHAPRCGIIIPASQVCGVYGFPSRVVVGTDHLAGPTSVRGENGTDLVLLNTPRPOGMMCTNMNSTGTTCCGGPPCNIGGGGNNLTCTPDCPKRHKPEATYTCGSGQPLPRLVNYPRILMYPCPTFNFTIKVRYVGVGHEHRSVACMWTPEBRCNLEDRRSELSPLLSITTRQIICGFTTLPALSTGLIHGQGIYVQYLYGICSAFVSFAIKWEYVLLLELLADRVCCWMLLQAQEALENLVNLAASVAGAHGILSFVFCAMVYIKGVLAGVAAYALGVPLLLALLPRAVAMDRMAASCGAVELISLSPYKVFARLIMWLOYETIRBAEHLVWVPLNVNGRADAILLTCAIHPLEIDITLLAILGLPLVWLAGTAVPYFAVRAQGLIRACMLRVKXVAGHYVOMALVKLAALGTGYVYDHLPIODMAHVADLAVAYEPVSDMETKLTITTCGAPACGDIILGLPSARSGKELLGRLPGLSGLGRLAPI TAYSOQTRGLACITITSLTRDKNOVEGEVOVSTATQSGSLCTINGCVMTYHGAGS KTLAAGKPIPTQMTYNVDOLGWLAPPGARSLPTCTGSSDLVYTRHADYIPVRR GDGRSLSPRVSYLKSGSGGLPCPSGVVGIIRAVACIGVAKADVAVESMET TMRSPVFTDNSSPPAVPOTFOYVHLHAPTSGSKRVPAAYAGGVKVLVNPVSAAT LSEFAYMSKAGHTDNIRGTGRTITTTGAPITTYSTYKGLADGCGGAGVADITCDECH STDSTTIIIGIGVLDQAEAGARLVVLTATPPGVYVPHPIEYVALSNTGEEVPIG KAIPIEATIRGGHLLIFCHSKKCDLAKLSALGNAAVYKGLDVSVIPNGDVVYV ADALMTGFTGDSVIDCNCVTQTVDFSLDPTPTITTTVPDVAVSQSGRTGTR GRRGIYREVTPEBPSGMESSVLCCEYDAGCAWELIPAEISILRAVYLNTPGPV ODHLEFMEGVFTGLTHIDAHFLSOTKQAGDNFVYAVATYCAARQAPPPSDMMK CLIRKPTLHGPTPLLYRLGANONEIILTHPTKTRIMCMSADLEVMVSTWVTCVGL ALAAYCITGSGSVYVGRILISGRPAVVPDEVLVTRPDEMECHSHPIYBOGQOLA BOFKOKALGLDITATKQAEAAVAVESKQALPFAVFAHMMWFISGIDYAGLALPLPG NPALASMAFASITSPITQSTLLFNILGVAQAOLAPPAVNSFAGLAGAAGVVS IGLRVLVDIILAGYAGVAGALVAFKVMGSEPPSTEDVNLPLAISGALVAGVCA ALIRRHVPGEGEAVHMMNRLLAFARSGNHVSPTHHVPSDAPALATLISNTITLOL KLRHOMINEDCSTPCSGMLRDWMDICTVLTDEMTYOSKTLPLRLPVPFPSCORGY KGVNRGESIMOTTCGCCNOIAGHYKNGSMRYVPTCTGSMVHGPIPIAATYTGPTPS PAPISTKALMVAEEIVEYTRVGDHFHTVGTQDITDNVCPQVPAPEFTVEDGRHL RYAPACKALDEVSFOVQLQYIVGSQLPEPEBDVAVILMSLMDPSHIAETARRR LARSPPLASSASOLSAPSLAKTCTHSDPADLIEANLIMQEGMGNTRRESE NKVYILDSFDPLRAEDEREVSVAPEILKRSKGPAPAIPIVARPNYNPPLLESKWD YVPPVYHGGCLPTRAPPIPPRRKRTVLTESYTSALATKATCTSSSSSANDSC TATAPPDASDDGKESDYESSMPLEGEBGDDLDGSKSTYSEASDEVCSSM STYTGALITVPCAAESKLPINALSNSILRNHNVLVATYTSASALRQKRVFEDRLQVL DDHRDVLKEMKAKASTVAKKLLSVEACKLTPPHSAKSKFGYAKGVNLSRAVNH

FEATURES

source

1. 10435

/organism="Rickettsia conorii"

IRSVKDLLEDITVTPIDITTVMAKNEVFCVOPKRGKRPALIVPFDLGVCEKMAKY DVAESTLPQAVMGSSYGFQYSPGQVPEFLVNAKSKCPMGFSDTSCPSVTENDIR VESISYOCDDLSPEAROAKSLKSLTERLYIGPLTNSKGNOCYRRRACSGVLTSCGNT LTCYKASACRAAKLODCMVLNDDILVITCDNSGVOEDACILAFIPAMRYSNAP GDPPOPAIDLEITSCSSNVSAHDSGRVYLTTRDPTPLARAMEYARTTPVNSP LGNITIMYAPTLWVMMI LMTWFFSILLAEQLEKALDCOYIGACYSIEPLDPIELR LHLASFSLSHSYSPGEINRYASCLARKLGVPRLTWHRRARSVRKALSGGAAATCGY LFNMAVTKTKLKLPIPAASOLDSCWFEVAGYSGDPIYHSLSRARRMFWLCLLSVGVGVLLPNR"

9378. 9415

/note="upstream of poly-U stretch"

9416. 9504

/note="poly(U) length is indistinct poly(U) stretch"

BASE COUNT 1886 a 2810 c 2683 g 2125 t

ORIGIN

Query Match 2.1%; Score 20; DB 14; Length 9504;

Best Local Similarity 100.0%; Pred. No. 22;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 311 CAGAAACACACAGAAAGA 330

|||||

Db 2662 CAGAAACACACAGAAAGA 2643

RESULT 48

AE008634/c

LOCUS Rickettsia conorii Malish 7, section 66 of 114 of the complete genome.

DEFINITION AE008634 AE006914

ACCESSION AE008634

KEYWORD AE008634.1 GI:15619870

SOURCE Rickettsia conorii.

ORGANISM Rickettsia conorii.

REFERENCE 1 (sites)

AUTHORS Ogata, H., Audic, S., Barbe, V., Artiguenave, F., Fournier, P. E., Raoult, D. and Claverie, J. M.

JOURNAL Selfish DNA in protein-coding genes of Rickettsia science 290 (5490), 347-350 (2000)

PMID 20485642

REFERENCE 2 (sites)

AUTHORS Ogata, H., Audic, S. and Claverie, J. M.

JOURNAL Selfish DNA and the origin of genes science 291 (5502), 252-253 (2001)

PMID 11030655

REFERENCE 3 (bases 1 to 10435)

AUTHORS Ogata, H., Audic, S., Renesto-Audiffren, P., Fournier, P. E., Barbe, V., Samson, D., Roux, V., Cossart, P., Weissenbach, J., Claverie, J. M. and Raoult, D.

JOURNAL Mechanisms of evolution in Rickettsia conorii and R. prowazekii science 293 (5537), 2093-2098 (2001)

PMID 11557893

REFERENCE 4 (bases 1 to 10435)

AUTHORS Ogata, H., Audic, S., Renesto-Audiffren, P., Fournier, P. E., Barbe, V., Samson, D., Roux, V., Cossart, P., Weissenbach, J., Claverie, J. M. and Raoult, D.

JOURNAL Direct Submission

Submitted (26-APR-2001) Unite des Rickettsies, Faculte de Medecine, CNRS UMR 6020, Universite de la Mediterranee, 27 Boulevard Jean Moulin, Marseille Cedex 05 13385, France

COMMENT A public version of R. conorii genome database is accessible at <http://igs-server.cns-mrs.fr/>. The database intends to provide updated data. Annotation of the genome is an ongoing task whose goal is to make the genome sequence more useful. Comments to the authors are appreciated

FEATURES

source

1. 10435

/organism="Rickettsia conorii"

gene /strain="Malish 7"
/db_xref="taxon:781"
complement(1170. .652)
/gene="RC0774"
complement(1170. .652)
/gene="RC0774"
/codon_start=1
/transl_table=11
/product="similarity to RP534"
/protein_id="AA03312.1"
/db_xref="GI:15619871"
/translation="MDFEADQKYSNYLHIGYCOTREHNNHDLGKADKIPDQREL
LNVEYVNKKLHHTFGFKLEEMFIOSEKNNTSKMGCVLENHDDKAKMLLSCCKGA
LNKEELNKLEKVIYIAATSRNNLTLMADTSLNALTLTLDVTASSKIKRSFIVYLE
"

gene complement(714. .1334)
/gene="RC0775"
complement(714. .1334)
/gene="RC0775"
/codon_start=1
/transl_table=11
/product="similarity to RP534"
/protein_id="AA03313.1"
/db_xref="GI:15619872"
/translation="MLKNNTGNFLQSRDIVNKKDAALDERRYNQDKITFYDII
ALRRYVNVQKDLVIMATQOETSELITFNSFDPNVEILACASASIPLVKPEVI
DSKATIDRRYGRNIPTEKFGNELEFDKATIDNMEETLAKKGGRIIAMAFTGMGT
DANIVYAKKRPESPDIIKFLADILFKTLAKVCGKFRYTENSKSN"
complement(1318. .1593)
/gene="RC0776"
complement(1318. .1593)
/gene="RC0776"
/codon_start=1
/transl_table=11
/product="similarity to RP534"
/protein_id="AA03314.1"
/db_xref="GI:15619873"
/translation="MEELREQYIPELMRQOKLVFYIMSRGCRFFRWTCYNCHSGSS
WPTRLSDLKFLPNLOTLGKKGFSADIIQLNRKDGKPIYALLEVIKE"
complement(1633. .1794)
/gene="RC0777"
complement(1633. .1794)
/gene="RC0777"
/codon_start=1
/transl_table=11
/product="similarity to RP534"
/protein_id="AA03315.1"
/db_xref="GI:15619874"
/translation="MADIKVRKRPILNNLKQESQPGVNEQSILIELLDLKKFNNLL
AIKQKHCRS"
complement(1799. .2428)
/gene="sodB"
/note="RC0778"
complement(1799. .2428)
/gene="sodB"
/codon_start=1
/transl_table=11
/product="superoxide dismutase [EC:1.15.1.1]"
/protein_id="AA03316.1"
/db_xref="GI:15619875"
/translation="MTYCNKSNQTSYPIPLDLYPEKSFPHFTLETFDYHGHKHNHN
TYVQNNLNLKDELOKDLDEIIEWSSONSNAIIPNNAOIVNHTFFWMSIPDGGSGT
GRSGKLTLOLNKDGSPERCEKQAEVGFSGMANLIYHDBRLDIIITANAQPIPI
IAHGKRPFLACDWEHAYITIDYRNNKRPDVVDLFIKHMINKFVEDNLK"
complement(2505. .2526)
/note="REP01, repeated element"
complement(2538. .2589)
/note="REP01, repeated element"
complement(2599. .3879)
/gene="folC"
/note="RC0779"
complement(2599. .3879)
/gene="folC"

CDS

/codon_start=1
 /product="folylpolyglutamate synthase [EC:6.3.2.11]"
 /transl_table=11
 /protein_id="AA03317.1"
 /db_xref="gi:15619876"
 /translation="MLMPHFVPPTMKNNIKYDLENIARILRALGNSHLRPPVHIHAG
 TNKSSAMSKSIFITLAGYVHCYTSPIHLENERIVAGKRIISDNELMAYCEQAV
 ASKREKVEPSEFGETTAAPLAFERKADILILETGGRDATNIVQPLITLPII
 SYDHMMALONTPLIATVYKAKIMKSCVCPKISMOYLEYENLFMAEELVPSVCY
 DEIKRTMGDTITYSSONTTEFPITPSLIGHQILNAAVIALISLKNQFVINSVEIA
 KGQNTIWPARELERKEPKYSKILGDVQWQIWDVAHNNAGVAILNMRDLKSLYL
 ILGNTNRIENIFECFEPFDILIKGVYKVLSEPLSHASARTINLEGRSGIDFSQDLSL
 EEAISDIKINDNNKANIITIGSDIYIASDPYKLL"
 3993..4021
 /note="REP02, RS3-like repeat"
 /complement(4066..4092)
 /note="REP02, RS3-like repeat"
 /complement(4132..4674)
 /gene="bioy"
 /note="RC0780"
 /complement(4132..4674)
 /gene="bioy"
 /codon_start=1
 /transl_table=11
 /product="similar to Biot protein"
 /protein_id="AA03318.1"
 /db_xref="gi:15619877"
 /translation="MQLNISLPLPTROAIEIVLGVALLTICQIVIPKIPFISLOT
 VAVIFGLITNATLPLTIFSLFITLGMAGIPVGEFSSGRIILFPTGGYIAGVIVIA
 YWASIKKDIPTNSQNMNOCITSLICGNIIIMSLGMMMLSLTLGASDAFYGVLPFIIP
 GIIKSVLLIGLINAVKPKTR"
 4750..5502
 /note="RC0781"
 /gene="RC0781"
 /db_xref="RC0781"
 /codon_start=1
 /transl_table=11
 /product="similarity to N-terminal of biotin-protein
 ligase"
 /protein_id="AA03319.1"
 /db_xref="gi:15619878"
 /translation="MKIKIYDLGYSKESIKHCYHPLRLTPAKRYNVDYITPAQELIDERN
 WYONFTLILIGSRDLTYVOKLOGGNANIKNYIKNGNFIIGCAGSYSGNVPFANIA
 GNEVYLSAREIKENGIVRGKPLAPYCNVSHKGARAYLKIPTLNIKCVAFYVY
 GGGFTIDAEYTKDTILASIEDSQAAIIKCYVGDGTAIIISGVILEYEPALLKNSLNN
 IHKIADHTERKLTNLTIDPFQNTLIERN"
 5463..5611
 /note="REP05, Rickettsia palindromic element 1"
 /complement(5614..6282)
 /gene="RC0782"
 /complement(5614..6282)
 /gene="RC0782"
 /codon_start=1
 /transl_table=11
 /product="similarity to glutathione S-transferase
 [EC:2.5.1.18]"
 /protein_id="AA03320.1"
 /db_xref="gi:15619879"
 /translation="MKKLIYHNPICGLSROARVPLKEIDIEFTIKEDKQNKDFLAK
 NPATPLTILIESGLFVVGITPFVYLNASKIPNPNLEIDEDIDICCEIRLPFPWFMDNIA
 VREYTKIILIDEKVIIRIAMKSSPTEFLRAKNLNLHLEITSLEKSYIYSSLSL
 IADDAACHSLVIDYGEIYMDKMTLLIKHWYSILKSRPARLLDRIIPGFPPNRYVY
 LIDP"
 6303..6322
 /note="REP03, repeated element"
 /complement(6360..6385)
 /note="REP03, repeated element"
 /complement(6394..6851)
 /gene="rnpB"
 /note="RCRNA21"
 /complement(6394..6851)
 /gene="rnpB"
 /product="RNA subunit (M1 RNA) of ribonuclease P"

repeat_region	complement(6624..6734)	/note="REP04, repeated element"
repeat_region	complement(6952..7025)	/note="REP06, A and T rich repeated element"
gene	7199..9847	/gene="ppdk"
CDS	7199..9847	/note="RC0783"
		/gene="ppdk"
Query Match	2.1%; Score 20; DB 1; Length 10435;	
Best Local Similarity	100.0%; Pred. No. 22;	
Matches	20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
OY	462 AAAAGTTATAGATCAATTG 481	
Db	5944 AAAAGTTATAGATCAATTG 5925	
RESULT 49		
AE013181/c	13610 bp DNA linear BCT 09-MAY-2002	
LOCUS		
DEFINITION	Thermoaerobacter tengcongensis strain MB4T, section 208 of 244 of the complete genome.	
ACCESSION	AE013181 AE008691	
VERSION	AE013181.1 GI:20517393	
KEYWORDS		
SOURCE	Thermoaerobacter tengcongensis.	
ORGANISM	Thermoaerobacter tengcongensis	
	Bacteria: Firmicutes; Clostridia; Thermoaerobacteriales;	
	Thermoaerobacteriaceae; Thermoaerobacter.	
REFERENCE	1 (bases 1 to 13610)	
AUTHORS	Bao,Q., Tian,Y., Li,W., Xu,Z., Xuan,Z., Hu,S., Dong,W., Yang,J., Chen,Y., Xue,Y., Xu,Y., Lai,X., Huang,L., Dong,X., Ma,Y., Ling,L., Tan,H., Chen,R., Wang,J., Yu,J. and Yang,H.	
	A Complete Sequence of the T. tengcongensis Genome	
	Genome Res. 12 (5), 689-700 (2002)	
TITLE	21992816	
JOURNAL		
MEDLINE	11997336	
PUBMED	2 (bases 1 to 13610)	
REFERENCE	Bao,Q., Xu,Z., Hu,S., Dong,W., Chen,Y., Wang,J., Yu,J. and Yang,H.	
AUTHORS		
TITLE	Direct Submission	
JOURNAL	Submitted (07-OCT-2001) Beijing Genomics Institute/Genomics and Bioinformatics Center, Institute of Genetics and Development, Chinese Academy of Sciences, Beijing Airport Industrial Zone B6, Beijing 101300, China	
	3 (bases 1 to 13610)	
REFERENCE	Li,W., Xuan,Z., Yang,J., Ling,L. and Chen,R.	
AUTHORS		
TITLE	Direct Submission	
JOURNAL	Submitted (07-OCT-2001) Institute of Biophysics, Chinese Academy of Sciences, Beijing 100101, China	
	4 (bases 1 to 13610)	
REFERENCE	Tian,Y., Xue,Y., Xu,Y., Lai,X., Huang,L., Dong,X., Ma,Y. and Tan,H.	
AUTHORS		
TITLE	Direct Submission	
JOURNAL	Submitted (07-OCT-2001) Institute of Microbiology, Chinese Academy of Sciences, Beijing 100080, China	
	Location/Qualifiers	
	1..13610	
FEATURES		
source	/organism="Thermoaerobacter tengcongensis"	
	/strain="MB4T"	
	/db_xref="taxon:119072"	
gene	complement(71..485)	
	/gene="G10A3"	
	/note="TTE2386"	
CDS	complement(71..472)	
	/gene="G10A3"	
	/note="Best Blastp hit"	
	g114010616 gb IAKS2053.1 AF364548_1 (AF364548)	
	methylmalonyl-CoA epimerase [Pyrococcus horikoshii], score	
	167, E-value 7.00E-41"	
	/codon_start=1	
	/transl_table=11	
	/product="Lactoylglutathione lyase and related lyases"	

```

misc_feature
|protein_id="AA025525.1"
/db_xref="GI:20517394.1"
/translacion="MIKKVDHIGIAVKSIEPAKFFTEYVGLKATGKIEEYKQKVKTA
FIPVSESSIELLESTSEDEGVPKAFTEKRGSEIHHIALQVDDIEKALEELKGGKGLD
EVPKRGAGAGIAFHPKPSANGVLELCEERK"
complement(74.442)
/gene="G10A3"
/note="Pfam match to entry Glyoxalase,
glyoxalase/Bleomycin resistance protein/Dioxygenase
superfamily, score 84, E-value 5.50E-23"
complement(78.472)
/note="TLR434, identity:97%, copy 2"
complement(474.1435)
/gene="ArgK"
/note="TTE2387"
complement(474.1418)
/gene="ArgK"
/note="Best Blastp hit = g1|7519922|pir1|F75215 Lysine
arginine ornithine transport system kinase PAB2398 -
Pyrococcus abyssi (strain Otsa)
g1|5457682|emb|CAB49173.1| (AJ248283) Lysine Arginine
ornithine TRANSPORT SYSTEM KINASE (Pyrococcus abyssi),
score 353, E-value 1.00E-96"
/codon_start=1
/transl_table=1
/product="putative periplasmic protein kinase ArgK and
related Gprases of G3E family"
/protein_id="AA025526.1"
/db_xref="GI:20517395"
/translacion="MMLEKELLNRKRAIRATTYAEDDEKAREIIRKLKRRSGNA
TIVATGGPVGKSLTDRLVKNLEBESKIGITAVDPTSEFTGSLGDIRQDLIT
THPVTISMGKRGHGLGSKGYMAVRIIDIAQMDYEVETGVGGSEVDIKIATP
VYVAVAPGIDGDIQAIKAGIMEIADVFNVNSGDEGADKYLENNMLDIDMPKEMP
PVYKVADEGKIGICEVLEKISHKHEYLLSSGLEERRNRIRKMETESVKNLKKDVPF
AKMGNEEDOKKQWAKEIDPYTAAELYHQILERR"
complement(481.485)
/gene="G10A3"
complement(1332.1822)
/note="TLR429, identity:95%, copy 2"
complement(1415.1836)
/gene="Sbm4"
/note="TTE2388"
complement(1415.1822)
/gene="Sbm4"
/note="Best Blastp hit = g1|6137077|emb|CAB59633.1|
(AJ246005) isobutyryl-CoA mutase, small subunit
[Streptomyces cinnamonensis]', score 195, E-value
2.00E-49"
/codon_start=1
/transl_table=1
/product="Methylmalonyl-CoA mutase, C-terminal
domain/subunit (cobalamin-binding)"
/protein_id="AA025527.1"
/db_xref="GI:20517396"
/translacion="MERPIRVLVAKPGIDGHDGKAVYATLRDAGMEVITYTGLRTP
EQIVAAIOEDVDVALSILSGAHNTLFPKIKTLKRGAGEDILVIGGVIAPEDDIPF
LKEHIAIEFTPGPTSAVIEYIKANVRLAKK"
complement(1431.1435)
/gene="ArgK"
complement(1454.1810)
/gene="Sbm4"
/note="Pfam match to entry B12-binding, B12 binding
domain, score 167.6, E-value 2.10E-46"
complement(1832.1836)
/gene="Sbm4"
complement(1840.3355)
/gene="Sbm5"
/note="TTE2389"
complement(1840.3519)
/gene="Sbm5"
/note="Best Blastp hit = g1|7519968|pir1|C75130
methylmalonyl-CoA mutase (EC 5.4.99.2) chain A
[stimulability] - Pyrococcus abyssi (strain Otsa)"

```



```
gene
/translation="MKFTFELLVLITLTVLSVSGVSAENDSTXPNXPKSNAPHEIIG
SMWIGRGRHNRNMKSSQGXHXSDXPOQXXXXDOK"
complement(join(4445..4573,5010..5139,5190..5275,
6469..6587,6740..6846,7023..7184,7843..7973,8258..8480,
9088..9188))
/gene="B0413.2"
CDS
complement(join(4445..4573,5010..5139,5190..5275,
6469..6587,6740..6846,7023..7184,7843..7973,8258..8480,
9088..9188))
/gene="B0413.2"
/notes="similar to 4 transmembrane segments integral
membrane proteins"
/codon_start=1
/protein_id="CAB07307.1"
/db_xref="GI:4008333"
/translation="MISCVACRQVAVGKRYKSNXXSNOTRYRLKPFEPFPLEFOVET
QCFEKTAKSLFLNKKCTKCTROFTGIRPLFLPFDLIQKLNSTFTFTTEKFKSL
QICKFEKXATSKIRGXVXXVXXVSPXSHNGXSVSDMTYTTDNTVNDSEVAN
RAGCGVWARYGIFTANIVFLIVGLXLAGVWLKIDSRENFISERYQAVOAEWEA
PTXFAVSTYXIVLGAVMVAVMAGCCGIGRSRPLIISMVVFLILVATLSGILYL
KKDGETVPAPISEFDRIXDPVIXKXDCGHRIMIALIGFSVTLIVESAVLQOV
LITCFALYFVFMREKEVKLIIVEXPHRHSRLTRDTLRDHLDFQLQKKRRRVASK
ER"
join(13685..13798,14087..14261,14308..14449,15299..15448,
15497..15760,16227..16332)
/gene="B0413.3"
join(13685..13798,14087..14261,14308..14449,15299..15448,
15497..15760,16227..16332)
/gene="B0413.3"
/notes="predicted using GeneFinder"
/codon_start=1
/protein_id="CAB07304.1"
/db_xref="GI:3873780"
/translation="MCOFTRIAPRLKCKANARBMHRNAXFASRVCXHFHKKTTISVT
LSTPKTKOSELISLSPAKPYMLAIGDSYSSGOCNPDIPTHSPRAVWDSCXSS
SNAFPLVSQKTXXTSSLSLJSCSGSTVDHGLISNGMREFFVSHIXAVARLE
ELNITRVXILDYDXTNRKXGVDSGAFGQACEXXFXILIXXIIIFXXKRIQKX
XAGLIRFELSRSPYOISLXKXKXKXKXKXKXKXKXKXKXKXKXKXKXKXKXKX
KXSLIRFELSRSPYOISLXKXKXKXKXKXKXKXKXKXKXKXKXKXKXKXKXKX
complement(join(16581..16640,16698..16822,17910..18332,
18748..18830,18902..19051,19138..19232,19383..19445,
19632..19732,20977..21018,21095..21248))
/gene="B0413.4"
complement(join(16581..16640,16698..16822,17910..18332,
18748..18830,18902..19051,19138..19232,19383..19445,
19632..19732,20977..21018,21095..21248))
/gene="B0413.4"
/codon_start=1
/protein_id="CAB07308.1"
/db_xref="GI:4008334"
/translation="MSDGNRRCFDFELRFGFVWKIMONTNYSXFLIPFSLILLI
XGCTYXAXASAKIYLSNCKNFQKICDGRPLIDENLSIFSORCGPDPFENOLC
PMTTALSSINRAAGKPNLXRSRTRAYVFOLESXELRTLAXEYPAKLCNKDTS
OPIAKWYFDITLSCLPFGCCEDDPDLKALRTGEEBENIEXKAPKXKXKXKX
LHMSAFAGRSYPOAIDVIOEKHIRGXAIEPTDVLVIVHKSDFAXXNLEX
SITVDEIXEKSENLAIEDNLLSTIMEGATVPXTSNLILEIDEXXEXGXSESL
EKSGNGEKXENRKRKXKXKXSEHSFEXPORNKSESCATAYRLCSCGASQSF
VYRMKEVGVCOSFPYGCXKQONVAHPRTRECBQXCD"
complement(19271..19342)
/gene="B0413.t2"
complement(19271..19342)
/gene="B0413.t2"
/notes="TTC Glu E-tRNA"
/product="tRNA-Glu"
predicted using tRNAscan-SE 1.11
preliminary prediction
similar to tRNA-Glu"
join(28590..28669,29866..30559)
/gene="B0413.5"
join(28590..28669,29866..30559)
/gene="B0413.5"
/codon_start=1
/protein_id="CAB07309.1"
```

```
/db_xref="GI:4008335"
/translation="MILXHVYRXHSFRTRRSHQKADFERFAXXPLASSXGXSXXI
PSLISPLIGFEGVYTAIOSTFGMKNRRIKGEIKDSGSDISDASEASDSECOIST
NFXOSISGRKLOKTKTENLEKSIIDYALCOKLPYLXLRXTXOIFRAOKXVYLR
IYXKXKXKXKXKXKXKXKXKXKXKXKXKXKXKXKXKXKXKXKXKXKXKXKXKX
XKXKXKXKXKXKXKXKXKXKXKXKXKXKXKXKXKXKXKXKXKXKXKXKXKX
KXKXKXKXKXKXKXKXKXKXKXKXKXKXKXKXKXKXKXKXKXKXKXKXKXKX
complement(join(32454..32734,33069..33211,33532..33701,
33775..33849,34037..34090))
/gene="B0413.6"
complement(join(32454..32734,33069..33211,33532..33701,
33775..33849,34037..34090))
/gene="B0413.6"
/notes="predicted using GeneFinder"
/codon_start=1
/protein_id="CAB07305.1"
/db_xref="GI:3873783"
/translation="MTNOKKVTREXIHREKSSVDAFYVHILRGKPYTKPLPLPTVQEA
DEPKXVYXSNHREPLRLSLTSDYQWDTNNRNRSSMTRLSFOFMDIPTXGP
LAXXILRRSKIRLREISEELRAMRIDMSLSTDWLOXOPFROHKYRLNLNEMAXX
QIDHLEQARAHLSRIINLETSPATPTEERKIDISWTENHEKLLAEMRAOAPPTHT
FOEXPTOTVVPVXXGXXXS"
complement(36074..36146)
/gene="B0413.t1"
complement(36074..36146)
/gene="B0413.t1"
/product="tRNA-Phe"
/notes="GAA Phe F-tRNA"
predicted using tRNAscan-SE 1.11
preliminary prediction
similar to tRNA-Phe"
join(37289..37544,39351..39461,39550..39627,39703..39726)
/gene="B0413.7"
join(37289..37544,39351..39461,39550..39627,39703..39726)
/gene="B0413.7"
/notes="cDNA EST yk481h5.5 comes from this gene
cDNA EST yk246b6.5 comes from this gene
cDNA EST EMBL:CO9106 comes from this gene
cDNA EST yk339g8.5 comes from this gene
cDNA EST yk440g8.5 comes from this gene"
/codon_start=1
/protein_id="CAB07310.1"
/db_xref="GI:4008336"
/translation="MESTOKDIKLEPDEFNYITAPGAPESSENTAENPENVDLSEXPL
EKRPAPFSSKGRSIVWEHXHTWAEORACNCAEPQVKGSTSLRLHRLKTIHPAA
INDDPNPPVARRKXKXKXKXKXKXKXKXKXKXKXKXKXKXKXKXKXKXKXKX
BASE COUNT 12836 a 6244 c 6214 g 12719 t 1713 others
ORIGIN
```

Search completed: April 26, 2003, 11:55:49
Job time : 3481 secs

THIS PAGE BLANK (USPTO)